Published in final edited form as:

J Mem Lang. 2013 October 1; 69(3): 349–367. doi:10.1016/j.jml.2013.05.002.

Limitations on adaptation to foreign accents

Alison M. Trude¹, Annie Tremblay², and Sarah Brown-Schmidt¹

¹Department of Psychology and Beckman Institute, University of Illinois at Urbana-Champaign

²Department of Linguistics, University of Kansas

Abstract

Although foreign accents can be highly dissimilar to native speech, existing research suggests that listeners readily adapt to foreign accents after minimal exposure. However, listeners often report difficulty understanding non-native accents, and the time-course and specificity of adaptation remain unclear. Across five experiments, we examined whether listeners could use a newly learned feature of a foreign accent to eliminate lexical competitors during online speech perception. Participants heard the speech of a native English speaker and a native speaker of Québec French who, in English, pronounces /i/ as [ɪ] (e.g., *weak* as *wick*) before all consonants except voiced fricatives. We examined whether listeners could learn to eliminate a shifted /i/-competitor (e.g., *weak*) when interpreting the accented talker produce an unshifted word (e.g., *weeze*). In four experiments, adaptation was strikingly limited, though improvement across the course of the experiment and with stimulus variations indicates learning was possible. In a fifth experiment, adaptation was not improved when a native English talker produced the critical vowel shift, demonstrating that the limitation is not simply due to the fact the accented talker was non-native. These findings suggest that although listeners can arrive at the correct interpretation of a foreign accent, this process can pose significant difficulty.

Keywords

speech; perception; eye-tracking; foreign accent; accommodation

When interpreting the speech of different talkers, listeners encounter a large amount of variability, including differences in pitch, speaking rate, and regional or foreign accent. Despite this variability, listeners normally adapt to these differences quickly, affording rapid and successful processing of the speech of various talkers. This ability to adapt to variability is surprising considering the many-to-many mapping problem in speech perception: Depending on the talker and phonetic context, a particular acoustic signal can map onto multiple phonemes, and a particular phoneme can be conveyed by the use of multiple acoustic signals (e.g., Gordon, 1988; Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967; Peterson & Barney, 1952).

Here we focus on a type of variability that poses a common, yet difficult challenge for listeners: the variability across talkers in regional and foreign accents. A central question is how listeners adjust to these talker- or dialect-specific changes in speech. In this paper, we

© 2013 Elsevier Inc. All rights reserved.

Contact Information: Alison M. Trude, 603 E. Daniel St., Champaign, IL 61820, (217) 244-4787, alison.trude@gmail.com.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

focus specifically on the case of a second-order constraint found in a foreign accent and address the mechanisms that allow listeners to tailor perception to a specific individual in multi-talker settings. Adaptation to foreign accents may confer high costs on listeners because of their phonetic and phonemic differences from native speech (e.g., Flege, 1987; Flege, Frieda, Walley, & Kandazza, 1998; Flege, Schirru, & MacKay, 2003; also see Flege, 1995). Such widespread phonetic and phonemic differences may be especially hard for listeners to overcome because they do not match the more typical acoustic characteristics associated with the speech of native talkers, and because they can create temporary or global ambiguities that would not exist in a native speaker's speech. For example, when speaking English, a native speaker of Québec French may produce the /i/ vowel as [ɪ] before all consonants except voiced fricatives due to a productive phonological process in Québec French. Thus, he or she may pronounce the word *weak* as [wɪ k], creating a lexical ambiguity with the word *wick*. Furthermore, native listeners may be less familiar with certain foreign accents than they are with regional dialects of their language, which may be encountered regularly in everyday life and through television or other media.

Despite these challenges, previous research has shown that although perception of foreign-accented speech is initially impaired, performance improves with minimal exposure, in some cases enough to equate performance with native speech (e.g., Bent & Bradlow, 2003; Bradlow & Bent, 2008; Clarke & Garrett, 2004; Sidaras, Alexander, & Nygaard, 2009; Weil, 2001). For example, Clarke and Garrett (2004) asked participants to complete a cross-modal matching task in which they listened to English sentences read by either a native or non-native speaker of English and responded to visual probes. After approximately 1 minute of exposure to the foreign accent, probe reaction times (which were initially slower with the accent) did not differ for native and non-native speech. Other research using transcription tasks suggests that when trained in multi-talker settings, learning generalizes to previously unheard talkers with the same accent, with comparable performance to listeners who were specifically trained on the test talker (Bradlow & Bent, 2008; also see Sidaras, Alexander, & Nygaard, 2009).

Although these studies demonstrate that listeners' comprehension of foreign-accented speech improves with exposure, questions about the mechanisms used to interpret this speech still remain. Specifically, because most of these studies used whole-sentence stimuli, listeners could have relied on sentence-level contextual information to guide interpretation (e.g., see Samuel, 1996). Thus, it is unclear how specifically listeners represented the phonetic features of the accent itself. Representing the features of a talker's accent could be important because knowledge about how a talker typically produces a given sound could allow for improvement in spoken word identification in cases where there is little context provided by sentence- or discourse-level information.

One exception to the existing line of research on accent adaptation is a study by Witteman, Weber, and McQueen (2013), in which native Dutch speakers completed a cross-modal priming task where the primes were single Dutch words produced by a native German speaker who spoke Dutch as a second language. Both Dutch and German have three diphthongs, but the diphthongs in the two languages differ. The critical stimuli were words in which the Dutch diphthongs were instead pronounced as German diphthongs. The results showed that the German-accented tokens primed the Dutch orthographic representations only for listeners who either had extensive prior experience with German-accented Dutch or were trained on other accented words containing the diphthongs at the start of the experiment. Further supporting the crucial role of listening experience in adaptation to accented speech, the authors found that listeners who were *not* trained on the critical diphthongs began to show priming effects in the second half of the experiment, presumably from exposure to the diphthong in the first half. These findings suggest that, even in an

impoverished, single-word context, listeners can learn a specific phonetic feature of a talker's accent and use that knowledge to affect their interpretation of the stimulus.

Another limitation of previous work in this area is that much of it uses off-line measures such as transcription and explicit judgments such as lexical decision. Thus, it is unknown whether the previously observed adaptations to foreign accented speech take place on-line as the speech is being perceived, or only in a delayed stage of processing. The answers to these questions will likely require the use of time-sensitive measures that tap into the online interpretation of foreign-accented speech, as well as methodologies that do not require the participant to make explicit judgments, allowing more insight into typical language processing in natural settings. The current work addresses these questions by using such an online measure, eye tracking in a visual world paradigm, and by testing participants' adaptation to one specific characteristic of the accent.

Some existing research suggests that listeners show rapid and specific learning and adaptation to accents, although in all of these cases, the accents used were from different regional dialects of the participants' native language rather than foreign-accented speech (e.g., Dahan, Drucker, & Scarborough, 2008; Evans & Iverson, 2003; Floccia, Goslin, Girard, & Konopczynski, 2006; Trude & Brown-Schmidt, 2012a; also see Maye, Aslin, & Tanenhaus, 2008 for similar evidence with synthesized accented speech). For example, Dahan, et al. (2008) exposed participants to a regional accent different from their own within the context of a visual-world eye-tracking task. The auditory stimuli were produced by a native English speaker with a regional accent in which the /æ/ vowel was raised to [e] before /g/ (e.g., *bag* [bæg]) but not before /k/ (e.g., *back* [bæk]). On critical trials, participants viewed a screen containing four orthographically-presented words, including an *-ag* word and an *-ack* word with the same initial consonant (e.g., *bag* and *back*). After exposure to the talker's accent, upon hearing the [bæ] of *back*, participants shifted their fixations to *back* and ruled out *bag* as a possible interpretation, demonstrating that participants had learned the phonetic characteristics of the talker's accent, ruling out *bag* as a competitor because it would have been pronounced with an [e] vowel by that talker. Subsequent research demonstrated that the same effect extends to multi-talker settings in which one of two randomly alternating talkers exhibits a different regional accent (Trude & Brown-Schmidt, 2012a), indicating that not only do participants adapt their phonetic representations to those of the current talker, but also they can quickly alternate their representations to guide on-line interpretation.

The Current Approach

Here we present the results of five experiments designed to test how listeners process non-native accents. Specifically, we examined whether listeners can rapidly learn a particular vowel shift present in the accent and apply this knowledge in time to affect online speech processing. As such, the present research adopts a visual world paradigm eye tracking task similar to the ones described above, allowing us to examine both learning and on-line processing. In each experiment, we employed a multi-talker design because of its similarity to real-life conversational situations in which several people take turns speaking, and because it allows us to test whether information about a particular talker can be retained and quickly applied during speech processing, even after exposure to an intervening talker (e.g., Kraljic & Samuel, 2007). We adopted a paradigm previously used with regional accents in which the critical feature of the accent involved a second-order constraint on the shift of a specific vowel (e.g., Dahan, et al., 2008; Trude & Brown-Schmidt, 2012a). In the present case, a native Québec French talker produced a shift while speaking English in which the /i/ vowel was lowered to [ɪ] if the consonant following it was *not* a voiced fricative (e.g., *wheeze* [wiz]; *weak* [wi k]). This talker did not begin learning English until age nine, and

therefore, her English exhibited other non-native-like features as well. Furthermore, her accent is one that our participants had never been exposed to. Thus, adapting to the accented speech in this study could potentially be much more difficult than in previous studies using this paradigm.

If listeners are able to rapidly learn and implement information about a talker's accent during on-line speech recognition, they should use knowledge of the accented talker's second-order vowel shift to eliminate potential accented competitors (e.g., *weak* [w₁ k]) when perceiving a cohort competitor without the vowel shift (e.g., *wheeze* [wiz]). Such a result would demonstrate that listeners can use specific information about the features of a foreign accent to guide online speech perception in a talker-dependent manner. By contrast, if listeners fail to learn the specific properties of this accent, when perceiving a word like *wheeze*, words like *weak* should continue to compete as they do with a native-English talker. Note that listeners may show some slight accommodation to the overall properties of the accent (e.g., Clarke & Garrett, 2004) as evidenced by faster lexical interpretation throughout the task, and an ability to use unique word-final information (e.g., -ze) to eventually identify the target within the context set. However, a failure to learn the specific vowel shift would be evident in a lack of an ability to rule out *weak*-like competitors, as listeners do with other types of accents (Dahan, et al., 2008; Trude & Brown-Schmidt, 2012a).

Across five experiments, we manipulate features of the training and testing set to examine characteristics of the learning and testing environment that might contribute to successful accommodation: In Experiments 1 and 2, all of the stimuli with shifted vowels were homophonous with other words in standard English pronunciation (e.g., *weak* was pronounced [w₁ k], which is also the pronunciation of the word *wick* in standard English). Success in these experiments would require listeners to learn that the Québec French talker produced these words as homophones because on some trials, the French talker produced [w₁ k] when naming a picture of *weak*, and on other trials, she produced the same sequence of phonemes when referring to a picture of *wick*. Experiment 1 included words in which the vowel shift was triggered by different coda consonants, whereas in Experiment 2, in an attempt to improve performance, we used only words in which the vowel shift was triggered by the coda consonant /t/. In Experiment 3, we somewhat alleviated the lexical ambiguity created by the vowel shift by removing trials in which the intended referent was the standard-pronunciation word (e.g., *wick*). In Experiment 4, we again attempted to improve performance, in this case by revising the stimulus set such that the accented pronunciation was not homophonous with another English word (e.g., *weed* was pronounced [w₁ d], a nonword in English). Finally, in Experiment 5, we manipulated whether a foreign-accented or native English speaker produced the critical vowel shift in order to examine whether the non-native status of the speaker affects the ability to learn this vowel shift. The results of these five experiments will not only reveal whether listeners can adapt to a specific second-order constraint of a foreign accent during on-line speech processing, but also the linguistic and contextual conditions under which they are more readily able to do so.

Experiment 1

The aim of Experiment 1 was to test whether native English-speaking participants could learn a specific phonological rule of a non-native accent (in this case, Québec French) and apply this information in a talker-specific manner during online speech processing. If listeners are successful, this would demonstrate that listeners encode specific phonological features of the accent and apply that information in time to affect on-line speech processing.

Native English-speaking participants listened to a native English speaker from the Chicago area and a native French speaker from Québec. The talkers alternated randomly across trials. Participants first listened to a recording of each talker reading a story so that they would be familiar with the speech of the two talkers. Participants then completed a visual-world eye-tracking task in which they viewed pictures on a computer screen and followed pre-recorded instructions from the English and French talkers while their eye movements were monitored.

Method

Participants—Sixty-two members of the University of Illinois community participated in Experiment 1. Three additional participants were excluded from analysis because the eye tracker could not be calibrated, and one additional participant was excluded because he knew the French talker. Participants received either \$16 or partial course credit. All participants were native speakers of American English and had normal or corrected-to-normal vision and hearing. The majority of participants spoke with a suburban Chicago accent, and due to the ubiquity of this accent among University of Illinois students, those who did not have it were certainly familiar with it. None of the participants reported having regular contact with anyone from Québec.

Stimuli—The stimuli were produced by a male and female talker who did not interact with the participants. The female talker was a native French speaker from a monolingual French-speaking town in Québec and began learning English in a classroom setting at nine years of age. Although her speech was accented, she did not consistently produce the critical vowel shift of interest; she was therefore instructed to produce the shift during the recording of the experimental stimuli. Importantly, the female talker was familiar with the vowel shift and easily produced the accented materials. The male talker was a native English speaker from the Chicago suburbs who did not produce this vowel shift. For convenience, we will refer to the female talker as the “French” talker and the male talker as the “English” talker, even though all stimuli were English words.

Participants listened to the French and English talkers during a training phase and a test phase. The stimuli for the test phase consisted of ten sets of seven monosyllabic English words (70 total words). For convenience, we will refer to words unaffected by the critical vowel shift as “unshifted” words, and those affected by the vowel shift as “shifted” words, though all the words produced by the French talker were clearly accented. Each stimulus set contained two critical words which shared an onset: an unshifted /i/-word (e.g., *wheeze*) and a shifted /i/-word (e.g., *weak*), with the latter being produced by the French talker as an /i/ (as in *wick*) and by the English talker as an /i/ (as in *weak*). Additionally, each set contained an unshifted /i/-word filler that shared an onset and coda with the corresponding shifted /i/-word in the set (e.g., *wick*, which was pronounced the same as *weak* by the French talker), as well as two pairs of unrelated filler words which shared an onset and a coda (e.g., *log* and *leg*) and contained vowels other than /i/ and /i/ (see Appendix A for a full list of stimuli).

In order to ascertain whether the French talker indeed produced the [ɪ] vowels similarly in the shifted words and their unshifted /i/-filler counterparts (e.g., *weak* and *wick*), as well as whether the [ɪ] vowels in these two word types indeed differed from the [i] vowels in the unshifted /i/-words, the average F1, F2, and duration were calculated for each stimulus word. The results, shown in Table 1 (means and 95% confidence intervals), indicate that the acoustic properties of the [ɪ] vowels in the shifted words and the unshifted /i/-fillers did not differ, and that the [ɪ] vowels did differ from the [i] vowels in the unshifted /i/-words. Conversely, the English talker’s [i] vowels in the unshifted and shifted /i/-words pattern together and are distinct from his productions of [ɪ] in the unshifted /i/-fillers. Note that

although the French talker's F1 values are higher than those of the English talker, crucially, her [i] vowels do not overlap with the English talker's [ɪ] vowels.

The acoustic stimuli for the training phase consisted of a story read individually by each of the two talkers. Each critical shifted and unshifted word was included in the story at least twice (with the exception of one word which was presented only once, see Appendix C for the full text). All acoustic stimuli were recorded in a sound-attenuated booth. All visual stimuli were color drawings taken from an online clipart database and were selected to provide the clearest possible depiction of the word.

Equipment and Procedure—Experiment 1 consisted of a training phase followed by a testing phase. The entire experiment lasted approximately 1 hour and 40 minutes. The experiment was programmed in Matlab using the Psychophysics Toolbox (PTB-3, Brainard, 1997; Pelli, 1997).

Training: Because of the phonemic constraints of our stimuli, some of the pictures used in the testing phase of the experiment may not have been easily identifiable due to low imageability (e.g., *fees*). Therefore, all participants first completed a picture training session. Participants viewed each picture along with an orthographic presentation of the picture name above the picture and a context-providing sentence below the picture. Each picture-name pair appeared once during the training, for a total of 70 trials. Following exposure, participants were tested on their knowledge of the picture names by viewing four of the experimental pictures on the screen, along with the name of one of those pictures (e.g., *fees*). Participants were asked to click on the picture that went with the name. Previous experiments demonstrated that one session of such training is generally successful (Trude & Brown-Schmidt, 2012a). Indeed, all participants successfully learned the pictures to a 90% or better criterion on their first attempt.

In the second phase of training, participants listened to a story read once by the English talker and once by the French talker. The story was designed to familiarize the participants with the two talkers' voices and expose them to each talker's pronunciation of the unshifted /i/-, unshifted /ɪ/-, and shifted /i/-words to be presented during testing. Participants always listened to the English talker first so that they would be familiar with the story's content when listening to the French talker. This minimized potential comprehension difficulties that may have arisen due to the French talker's accent and ensured that participants would be able to correctly understand the shifted /i/-words. Before hearing each story, participants were informed of the upcoming talker's country of origin (United States or Canada) and native language (English or French). No additional tasks were imposed upon the participants during this portion of the experiment.

Testing: During the testing phase, participants' eye movements were recorded using an Eyelink 1000 desktop-mounted eye tracker which sampled eye position monocularly at 1,000 Hz. Participants first viewed a fixation cross for 1,000 ms. Participants then viewed a display containing four pictures from one of the ten word sets. On critical trials, the display contained an unshifted /i/-word (e.g., *wheeze*), a shifted /i/-word (e.g., *weak*), and one of the two possible pairs of fillers (e.g., *log* and *leg*), as illustrated in Figure 1 (see Table 2 for a list of critical trial types).

A complementary set of trials included an unshifted /ɪ/-word filler rather than a shifted /i/-word. After 2,000 ms, the participants heard the preamble "*Select*," followed by 200 ms of silence, then followed by the target word, and were instructed to click on the word that they had heard. Participants were given unlimited time to respond. The preamble and target word were both spoken by the same talker. All audio was played through desktop speakers. The

test consisted of 560 trials (280 per talker). Each of the 70 words used in the experiment appeared as the target word four times per speaker. Note that the unshifted /i/-words appeared more frequently on the screen than any of the other words (because they served as competitors for both the unshifted /i/-word fillers and shifted /i/-words), but that these words were the auditory target equally frequently. However, note that over the course of the experiment, when talker identity is ignored, the backward probability of having heard /i/ when the target is an unshifted /i/-word is 100%, whereas the backward probability of having heard /i/ when the target is a shifted /i/-word is only 50% (on the English talker trials, but not the French talker trials). This imbalance could cause listeners to be less likely to fixate the image depicting the shifted /i/-word when the auditory stimulus contains an [i]. The trials were presented to each participant in a different random order to ensure that participants could not predict which talker was going to speak next.

Predictions—Our analysis focuses on trials with a potentially accented word in the display. If listeners can learn the accent, and use this knowledge to guide on-line processing on trials with an unshifted /i/-target (e.g., *wheeze*) and a shifted /i/-competitor (e.g., *weak*), listeners should show more fixations to the target when listening to the French talker than when listening to the English talker. This is because for the French talker, *wheeze* and *weak* contain different vowels (/i/ and /ɛ/, respectively), creating less competition than for the English talker, who pronounces both words with the same /i/ vowel. Likewise, when the shifted /i/-word (e.g., *weak*) is the target and the unshifted /i/-word (e.g., *wheeze*) is the competitor, we would predict a similar outcome, as the French talker's pronunciation of *weak* as [wɛ k] could match only the target.

A different pattern of results is expected if listeners fail to learn the specific vowel shift, or fail to use their knowledge of the accent during on-line processing. If the specific vowel shift is not learned, listeners may process the French talker's speech using standard representations of English (see Sumner & Samuel, 2009 for a related discussion). If so, we would predict similar patterns of fixations for both the French and English talkers in all conditions except the one in which the shifted /i/-word (e.g., *weak*) is the target. That is, when an unshifted word is the target, listeners should initially fixate both the target and competitor at equal rates for the French and English talkers. In contrast, when the shifted word is the target, a listener who had not learned the characteristics of the accent would likely be surprised by the unfamiliar accented word and may show delayed processing in comparison to the English talker.

Analysis and Results

Participants' target selection accuracy was above 95% for all trial types, indicating that listener's ultimate interpretation was accurate, despite the potential for homophony created by the accented word (note that the accented word *weak* never appeared on the screen with the homophonous *wick*). This result is consistent with previous findings of accommodation of foreign accents in off-line tasks (Bradlow & Bent, 2008), but does not speak to whether the online processing of these words was as rapid as with native speech, nor whether listeners used knowledge of the accent to eliminate temporary ambiguity in the acoustic signal. We now turn to an analysis of eye gaze during perception of the critical words to examine on-line processing.

The eye movement data were analyzed in seventeen 100 ms regions, from 200 to 1,600 ms relative to critical word onset. In addition, a baseline region from 0-200 ms following word onset was used to evaluate whether listeners had fixation biases prior to hearing the critical word. The critical region is offset by 200 ms due to the time needed to program and launch an eye movement (e.g., Hallet, 1986). The ratio of the number of fixations to the target over

the number of fixations to the competitor in each region was calculated separately for subjects and items. The natural log of this term was used as the dependent measure (for similar approaches, see Brown-Schmidt, 2012; Heller, Grodner, & Tanenhaus, 2008; Ferguson, Scheepers & Sanford, 2010); an equal preference would result in a score of zero, a preference for the target would result in a positive score, and a preference for the competitor, a negative score.¹ The advantage of using a log scale is that, unlike proportions, they are not bounded at 0 and 1. However, in order to more clearly highlight the effects of lexical competition, we present figures plotting the proportion of fixations to the target, competitor, and average of the two fillers. For each condition, repeated measures ANOVAs by subject and item were performed with talker (English vs. French) and time (200-1,600 ms in 100 ms intervals) included as within-subjects factors.

Unshifted /i/-word target (e.g., *wheeze*), shifted /i/-word competitor (e.g., *weak*)

—Figure 2 presents the proportions of fixations to target, competitor, and filler words when the target was an unshifted /i/-word (e.g., *wheeze*) and the competitor was a shifted /i/-word (e.g., *weak*). Analysis on the time region from 0-200 ms revealed no baseline effects ($t = .66$). The ANOVA on the critical region (200 to 1,600 ms post-onset) showed a main effect of time, $F(14, 854) = 167.2, p < .001, (\epsilon = .244)$; $F(14, 126) = 42.3, p < .001, (\epsilon = .139)$, due to an increase in the preference to fixate the target as the critical word unfolded. A main effect of talker also emerged, $F(1, 61) = 13.0, p = .001$; $F(1, 9) = 10.6, p = .01$, with participants fixating the target more on English talker than French talker trials ($M = 1.81$ and $M = 1.47$, respectively). These effects were qualified by a time-by-talker interaction, which was significant by subjects, $F(14, 854) = 4.35, p = .001, (\epsilon = .338)$; $F(14, 126) = 2.38, p = .12, (\epsilon = .157)$. Planned comparisons indicated that, late in the trial, participants fixated the target more when listening to the English talker. This effect was significant from 1,100-1,600 ms by subject ($ts > 1.98$) and from 1,300-1,500 ms by item ($ts > 2.52$). These results suggest that listeners were not more likely to eliminate the competitor when processing the French talker's speech. In fact, late in the trial, the opposite was true: participants fixated the target less when listening to the French talker.

Shifted /i/-word target (e.g., *weak*), unshifted /i/-word competitor (e.g., *wheeze*)

—Figure 3 presents the proportions of fixations to target, competitor, and filler words in the condition where the target was a shifted /i/-word (e.g., *weak*) and the competitor was an unshifted /i/-word (e.g., *wheeze*). Analysis of the time region from 0-200 ms indicated that there was no baseline difference between talker conditions ($t = .51$). The ANOVA on the critical region (200 to 1,600 ms post-onset) revealed main effects of both time $F(14, 854) = 222.8, p < .001, (\epsilon = .168)$; $F(14, 126) = 81.5, p < .001, (\epsilon = .141)$ and talker $F(1, 61) = 204.4, p < .001$; $F(1, 9) = 65.0, p < .001$. These effects were qualified by a significant interaction between talker and time, $F(14, 854) = 96.8, p < .001, (\epsilon = .269)$; $F(14, 126) = 29.4, p < .001, (\epsilon = .155)$. Contrary to what would be predicted under an adaptation account, participants fixated the target *less frequently* when listening to the French than the English talker ($M = .58$ and $M = 1.51$, respectively). Planned comparisons indicated that the source of the interaction was a crossover effect such that on French talker trials there was a significantly larger target advantage from 300 to 500 ms by subjects and 300 to 400 ms by items, and significantly smaller target advantage on French talker trials from 700 to 1,600 ms by subject ($ts > 3.54$) and from 800 to 1,600 ms by item ($ts > 4.89$).

The apparent initial advantage for the French talker was largely driven by participants' early preference for the unshifted /i/-word competitor (e.g., *wheeze*) on English talker trials. On

¹In cases where the # of fixations to the target or competitor was zero (in a given region), $\ln(\text{target}/\text{competitor})$ is undefined. In these cases, a single fixation was added to both numerator and denominator in the calculation of the target preference ratio, (see Heller, et al., 2008 for a related approach); this affected 5.8% of data points.

French talker trials, participants showed little preference for either image before 500 ms (note the closeness of the target and competitor lines in Figure 3), whereas they showed a preference for the *competitor* picture on English talker trials in that same time window (note that for the English talker, the competitor line is above the target line). This tendency to initially interpret the target word as an unshifted /i/-word when listening to the English talker may have been due to the previously-mentioned differences in backward probability present in our stimuli. Thus, participants may have adopted a strategy of fixating the unshifted /i/-word competitor when hearing the English talker produce an /i/ vowel. This competitor preference may not have been observed for the French talker trials due to participants being less certain overall when interpreting her speech, thus leading them to fixate both pictures roughly equally.

Learning across the experiment—In order to examine whether participants improved at interpreting the French talker's speech over the course of the testing phase, target preference scores in the first and last third of the experiment were calculated over a large time-window (200-1,600 ms). Repeated measures ANOVAs were performed by subject and item in each condition, with talker (French vs. English) and third (first vs. last) as within-subjects factors. Because we are concerned with change in performance over the course of test, we only focus on the effects of third and its potential interaction with the effect of talker (see Appendix B for descriptive statistics for all experiments). For trials with an unshifted /i/-target (e.g., *wheeze*), there was an effect of third that was significant by items, $F_{\chi}(1, 9) = 5.5, p < .05$, due to improvement with both talkers in the last third of the experiment. For trials with a shifted /i/-word target (e.g., *weak*), there was also a main effect of third significant by items, $F_{\chi}(1, 9) = 5.5, p < .05$. However, this was qualified by a talker-by-third interaction, ($F_f(1, 61) = 61.0, p = .01$; $F_{\chi}(1, 9) = 6.7, p < .05$), due to more improvement over time with the French than the English talker.

The fact that listeners improved over the course of the experiment is important because it shows that learning was, in fact, possible within our paradigm, despite the fact that listeners were not able to successfully use this learning to rule out potential competitors. The enhanced learning with the French talker is likely due to the fact that performance was overall worse with the French talker, particularly at the beginning of the experiment, and thus there was more room for improvement (i.e., a scaling effect). Error-driven learning (i.e., a larger error signal when initially failing to interpret the French talker) may have also contributed.

Discussion

When interpreting an unaccented word like *wheeze* in the context of a potentially accented word like *weak*, we hypothesized that if listeners were able to use their knowledge of the French talker's accent, they should fixate the target more when listening to the French talker than when listening to the English talker because the French talker produces these words with different vowels ([i] and [ɪ]). However, contrary to predictions, participants' performance was equivalent in the two talker conditions until 1,100 ms after target onset, at which point participants fixated the target *more* when listening to the English talker. Participants' representations of the French talker's speech apparently did not include the fact that she would have produced the shifted /i/-word with an [ɪ], and thus, competition was not attenuated. Additionally, participants' lower likelihood of fixating the target late in the trial when listening to the French talker indicates that listeners experienced residual uncertainty about the interpretation of the accented speech. This finding suggests that listeners did *not* use specific knowledge of the accent to guide on-line processing, unlike the interpretation strategy seen with regional accents (Dahan et al., 2008; Trude & Brown-Schmidt, 2012a). At the same time, the fact that participants improved over the course of the experiment is

consistent with previous findings of rapid adaptation to foreign-accented speech (e.g., Bent & Bradlow, 2003; Bradlow & Bent, 2008; Clarke & Garrett, 2004; Sidaras, Alexander, & Nygaard, 2009; Weil, 2001), and is likely facilitated by the fact that both talkers produce the unshifted /i/-words with the canonical /i/ vowel and the unshifted /ɪ/-word fillers with the canonical /ɪ/ vowel.

In contrast, when listeners were presented with a shifted /i/-word word as the target, we observed pronounced competition between the two words in the French talker condition that continued until at least 1,600 ms after word onset. The competition effects lasted well after the disambiguating coda consonant, suggesting that lexical competition between the target and competitor was largely unresolved.

In summary, while participants showed some evidence of accommodating the accent, allowing them to click the correct picture, listeners' online processing showed significant deficits. Listeners' failure to learn and implement knowledge about the accent comes as a surprise, given that the method of Experiment 1 closely resembled that of experiments by Dahan et al. (2008) and Trude and Brown-Schmidt (2012a), in which participants successfully used knowledge of a regional accent vowel shift to guide on-line processing. However, another possibility is that the differences in the stimuli between the previous work on regional accent adaptation and our Experiment 1 made learning of the accent harder for participants. In the regional accent studies, all accented words presented during the course of the experiment ended in /g/. Conversely, in Experiment 1, the shifted /i/-word stimuli ended in a variety of codas. It may be that consistently presenting the accented vowel followed by the same coda made it easier for listeners in the regional accent studies to notice and learn the context in which the accented vowel occurred. We explore this possibility in Experiment 2 by simplifying the set of shifted /i/-words such that they all share the same coda.

Experiment 2

Experiment 2 tested whether presenting participants with a more phonotactically constrained set of shifted /i/-words would improve their learning of the foreign accent. The same talkers from Experiment 1 produced the stimuli for Experiment 2, but all shifted /i/-words and unshifted /ɪ/-word counterparts were /t/-final. As in Experiment 1, all of the unshifted /i/-words had voiced fricative codas (/z/ and /v/). We predicted that if the constrained set of stimuli made the French talker's vowel shift easier to learn, we should observe a French-talker advantage. However, if participants have difficulty learning the accent even with the more limited set of stimuli, we may expect a pattern of results that looks similar to that found in Experiment 1.

Method

Participants—Sixty-three University of Illinois students who did not participate in Experiment 1 participated in Experiment 2. Five additional participants were excluded due to technical errors. Participants received partial course credit for participating. All participants were native speakers of American English and had normal or corrected-to-normal vision and hearing. None of the participants reported having regular contact with anyone from Québec.

Stimuli—The stimuli were produced by the same male native English speaker and female native French speaker from Experiment 1. The stimuli for the test phase consisted of seven sets of seven English words. As in Experiment 1, each word set contained two critical words which shared an onset: an unshifted /i/-word (e.g., *wheeze*) and a shifted /i/-word (e.g., *wheat*), which the French talker produced with an /ɪ/ (as in *wit*) and the English talker produced with an /i/. Unlike Experiment 1, all shifted /i/-words had a /t/ coda. The

unshifted /i/-words all had voiced fricative codas (/z/ and /v/). Additionally, each word set contained an unshifted /i/-word filler (which also had a /t/ coda) and two pairs of unrelated filler words. Each pair of unrelated filler words shared an onset and a coda (e.g., *leg* and *log*). The filler words all contained vowels other than /i/ and /ɪ/ (see Appendix A for a full list of the stimuli). Analysis of the vowels in the French talker's productions of the shifted /i/-words and unshifted /i/-filler counterparts (e.g., *wheat* and *wit*), shown in Table 3, ensured that the stimulus words were produced with the same vowel. Furthermore, the productions of the [i] vowels in these two word types differed in duration and F1 from the [i] vowels in the unshifted /i/-words. The 95% confidence interval for the F2 of the unshifted /i/-word vowels overlapped slightly with the confidence intervals for the two categories of [i] vowels; however, the large differences in the duration and F1 of the French talker's [i] and [ɪ] vowels suggest that she did systematically produce these two vowels differently, and that this contrast should have been perceptible to participants. As in Experiment 1, the English talker's [i] vowels in the unshifted- and shifted /i/-words were similar, and both differed from his production of /ɪ/ in the unshifted /i/-fillers. Again, the French talker's [i] vowels were produced with a higher F1 than the English talker's [i] vowels, but they did not overlap with his [ɪ] vowels.

Like in Experiment 1, the acoustic stimuli for the training phase consisted of a story read individually by each talker. The story was different than the one in Experiment 1 due to the new stimuli, and contained two instances of each critical unshifted /i/-, and shifted /i/-word (see Appendix C for the full text). All individual word stimuli were recorded in a sound-attenuated booth. The stories were recorded directly to disk using a headset microphone. Each talker was recorded separately. All visual stimuli were color drawings taken from an online clipart database and were selected to provide the clearest possible depiction of the associated word.

Equipment and Procedure—Like Experiment 1, Experiment 2 consisted of a training phase followed by a testing phase. The entire experiment lasted approximately 1 hour and 20 minutes.

Training: All participants first completed a picture training session that used a procedure identical to that used in Experiment 1. Each critical word and filler word was seen once during the training, for a total of 49 trials. All participants successfully learned the pictures to a 90% or better criterion on their first attempt. Next, participants listened to a story read once by the English talker and once by the French talker. As in Experiment 1, the participants always listened to the English talker read the story before the French talker.

Testing: The testing phase of Experiment 2 was the same as in Experiment 1. It consisted of 392 trials (196 per talker). Each of the 49 stimulus words was the target word four times per talker. The trials were presented to each participant in a different random order.

Results

The analyses for Experiment 2 were conducted identically to those for Experiment 1. Accuracy rates exceeded 97% for all trial types.

Unshifted /i/-word target (e.g., *wheeze*), shifted /i/-word competitor (e.g., *wheat*)—Figure 4 presents the proportions of fixations to target, competitor, and filler words in the condition where the target was an unshifted /i/-word (e.g., *wheeze*) and the competitor was a shifted /i/-word (e.g., *wheat*). Analysis from 0-200 ms revealed no baseline effects ($t = .81$). The ANOVA on the critical region (200 to 1,600 ms post-onset) showed a main effect of time, $F_1(14, 868) = 398.2, p < .001, (\eta^2 = .335)$; $F_2(14, 84) = 56.6, p < .001, (\eta^2 = .335)$.

= .147). The main effect of talker was significant by subjects and marginal by items, $F(1, 62) = 8.8, p < .01$; $F(1, 6) = 5.5, p = .06$, with participants fixating the target more when listening to the English talker ($M = 2.00$ and $M = 1.74$ for English and French talkers, respectively). Additionally, the talker-by-time interaction was significant by subjects, $F(14, 868) = 4.0, p < .01$, ($\epsilon = .382$); $F(14, 84) = 2.7, p = .10$, ($\epsilon = .167$). Planned comparisons showed significant differences between talker conditions from 800-1,600 ms by subjects ($ts > 1.99$) and from 900-1,200 ms by items (marginal at 1,000 ms, $p = .06$) ($ts > 2.39$), with listeners making *fewer* target fixations when listening to the French talker.

Shifted /i/-word target (e.g., *wheat*), unshifted /i/-word competitor (e.g., *wheeze*)—Figure 5 presents the proportions of fixations to target, competitor, and filler words in the condition where the target was a shifted /i/-word (e.g., *wheat*) and the competitor was an unshifted /i/-word (e.g., *wheeze*). Analysis of the time region from 0-200 ms indicated that there was no baseline effect ($t = .01$). The ANOVA on the critical region (200 to 1,600 ms) revealed main effects of time, $F(14, 868) = 239.9, p < .001$, ($\epsilon = .224$); $F(14, 84) = 93.8, p < .001$, ($\epsilon = .147$) and talker, $F(1, 62) = 673.1, p < .001$; $F(1, 6) = 111.3, p < .001$. These effects were qualified by a significant talker by time interaction $F(14, 868) = 86.1, p < .001$, ($\epsilon = .327$); $F(14, 84) = 47.9, p < .001$, ($\epsilon = .191$). Planned comparisons indicated that this difference was significant from 300-1,600 ms by subject ($ts > 2.03$; marginal at the 400 ms window, $p = .06$) and 500-1,000 ms by item ($ts > 2.76$; marginal at the 400 ms window, $p = .06$). This effect was due to a lower proportion of target fixations on French talker than English talker trials ($M = 2.42$ and $M = .49$, respectively).

Learning across the experiment—As in Experiment 1, we compared performance between the first and the last third of the experiment to test for learning; again we focus only on the effects of third. For unshifted /i/-words (e.g., *wheeze*), there was a talker-by-third interaction significant only by items, $F(1, 6) = 8.7, p < .05$, with performance slightly increasing on French talker trials and slightly decreasing on English talker trials in the last third of the experiment. By contrast, shifted /i/-words (e.g., *wheat*) showed the same interaction pattern as in Experiment 1. The main effect of third was significant, $F(1, 62) = 4.5, p = .05$; $F(1, 6) = 7.6, p < .05$, with improvement in the final third of the test. This effect was qualified by an interaction such that listeners improved more with the French than the English talker, $F(1, 62) = 6.4, p = .01$; $F(1, 6) = 5.4, p = .06$. This result again shows that listeners did learn, albeit slowly, about the French talker's vowel shift with increasing exposure.

Discussion

The results of Experiment 2 showed that despite simplifying the phonotactic regularity of the set of accented words, listeners did not take advantage of the characteristics of the French talker's accent to guide online processing. This finding demonstrates that the phonotactic regularity of the accented word set is *not* the critical feature that explains why listeners were able to use knowledge of a regional vowel-shift to guide on-line processing (Dahan et al., 2008; Trude & Brown-Schmidt, 2012a), but not knowledge of the formally equivalent vowel-shift in Experiments 1-2. Why, then, is interpretation of the French talker's speech particularly difficult? One possible explanation is that the homophony between all the shifted /i/-words (e.g., *wheat*) and the unshifted /i/-word fillers (e.g., *wit*) confused participants. Although these two word types were never presented at the same time, across trials the same sequence of phonemes could refer to either an unshifted /i/-word or a shifted /i/-word. Listeners may have struggled with this lexical ambiguity, searching for the image of *wit* when hearing the French talker produce [w_i t], even though the correct picture in the display was *wheat*.

In Experiment 3, we eliminated this source of difficulty by excluding the unshifted /i/-word fillers from the experiment. Thus, listeners could be certain that if they heard the French talker produce a word containing an [i] vowel, the referent would be the shifted /i/-word and not the homophonous unshifted /i/-word (e.g., [w₁ t] would always refer to *wheat*; never to *wit*).

Experiment 3

If lexical ambiguity due to homophonous stimuli was the source of listeners' difficulty, we predict that removing unshifted /i/-words from the stimulus set should support more robust learning. If so, on French-talker trials with an unshifted /i/-target (e.g., *wheeze*) and a shifted /i/-competitor (e.g., *wheat*), participants should fixate the target more with the French talker.

Method

Participants—One hundred and three University of Illinois students participated in Experiment 3. One additional participant was excluded because the eye tracker could not be calibrated, and one was excluded due to a technical error. Participants received partial course credit for participating. All were native speakers of American English and had normal or corrected-to-normal vision and hearing. No participant reported having regular contact with anyone from Québec; none had participated in Experiment 1 or 2.

Stimuli—Visual and auditory stimuli were the same as those used in Experiment 2 (see Appendix A); however, the unshifted /i/-words (e.g., *wit*) were not presented at all during the test. Nevertheless, the stories used in the training phase were the same as those used in Experiment 2; therefore, the unshifted /i/-words were present in the training.

Equipment and Procedure—Like Experiments 1 and 2, Experiment 3 consisted of a training phase followed by a testing phase. The entire experiment lasted approximately 60 minutes.

Training: All participants first completed a picture training session with a procedure identical to that used in Experiments 1 and 2. Each critical word and filler word was seen once during the training, for a total of 42 trials. All participants successfully learned the pictures to a 90% or better criterion on their first attempt. Next, participants listened to the same training stories as were used in Experiment 2. As in the previous experiments, the participants always listened to the English talker read the story before they listened to the French talker.

Testing: The testing phase was the same as in Experiments 1 and 2. It consisted of 196 trials (98 per talker). Each of the 42 stimulus words was the target word four times per talker, except the shifted /i/-words (e.g., *wheat*), which were the target eight times per talker. The trials were presented to each participant in a different random order.

Results

The results of Experiment 3 were subjected to the same analyses as those used in Experiments 1 and 2. Accuracy exceeded 98% in all conditions.

Unshifted /i/-word target (e.g., *wheeze*), shifted /i/-word competitor (e.g., *wheat*)—Figure 6 presents the proportions of fixations to target, competitor, and filler words when the target was an unshifted /i/-word (e.g., *wheeze*) and the competitor was a shifted /i/-word (e.g., *wheat*). Analysis from 0-200 ms after word onset indicated there were

no baseline differences between conditions ($t = .57$). The ANOVA at the critical region (200-1,600 ms) revealed an effect of time, $F(14, 1428) = 446.2, p < .001, (\epsilon = .322)$; $F(14, 84) = 59.2, p < .001, (\epsilon = .119)$. A main effect of talker was marginal by subjects, $F(1, 102) = 3.8, p = .053$, with more target fixations in the English talker than the French talker condition ($M = 1.82$ and $M = 1.66$, respectively). These effects were qualified by a talker-by-time interaction, $F(14, 1428) = 3.5, p < .01, (\epsilon = .335)$; $F(14, 84) = 3.2, p < .01, (\epsilon = .243)$. Planned comparisons indicated that the talker effect was significant from 1,100-1,600 ms by subject ($ts > 2.27$) and from 1,500-1,600 ms by item ($ts > 3.63$).

Shifted /i/-word target (e.g., *wheat*), unshifted /i/-word competitor (e.g., *wheeze*)—Figure 7 presents the proportions of fixations to target, competitor, and filler words in the condition where the target was a shifted /i/-word (e.g., *wheat*) and the competitor was an unshifted /i/-word (e.g., *wheeze*). A baseline analysis indicated no significant baseline difference between talker conditions ($t = 1.95$). The ANOVA at the critical region showed main effects of both time, $F(14, 1428) = 442.4, p < .001, (\epsilon = .311)$; $F(14, 84) = 106.9, p < .001, (\epsilon = .157)$, and talker, $F(1, 102) = 363.5, p < .001$; $F(1, 6) = 76.9, p < .001$. The talker effect was due to more target fixations on English talker trials ($M = 2.18$ and $M = .82$, respectively). Planned comparisons indicated the talker effect was significant from 500-1,600 ms by subject ($ts > 3.64$) and from 400-1,600ms by item ($ts > 2.91$). Both effects were qualified by a significant talker-by-time interaction, $F(14, 1428) = 84.6, p < .001, (\epsilon = .317)$; $F(14, 84) = 24.8, p < .001, (\epsilon = .183)$.

Learning across the experiment—As in Experiments 1-2, learning analyses compared performance in the first and final thirds of the experiment. There was no effect of learning on unshifted /i/-word (e.g., *wheeze*) trials. For shifted /i/-words (e.g., *wheat*), there was a main effect of third, $F(1, 102) = 14.6, p < .001$; $F(1, 6) = 5.8, p = .05$, with better performance in the last third of the test, whereas the by-item analysis also revealed a talker-by-third interaction, $F(1, 6) = 7.4, p < .05$, due to a greater increase in performance on the French than the English talker trials. These findings suggest listeners improved with both talkers over the course of the experiment.

Discussion

On critical trials in Experiment 3, we predicted that if listeners use their knowledge of the French-talker's accent, they should fixate the target more when listening to the French talker than the English talker. Contrary to our expectations, on trials with an unshifted /i/-word target (e.g., *wheeze*) and a shifted /i/-word competitor (e.g., *wheat*), we again found that participants' performance was better on English talker trials. This result suggests that listeners did not use cues to target identity provided by the French talker's accent.

One possible explanation for listeners' continued failure to use the vowel shift for ruling out lexical competitors is that although unshifted /i/-words (e.g., *wit*) did not appear in Experiment 3, participants still activated those words when hearing the French talker's pronunciations of shifted /i/-words (e.g., *wheat*). Thus, the homophony created by the French accent may have caused competition from those lexical items, potentially consistent with other evidence of competition from potential competitors that are present in the lexicon but not in the currently-viewed scene (Magnuson, Dixon, Tanenhaus, & Aslin, 2007).

In Experiment 4, we eliminated the possibility that homophonous unshifted /i/-words were activated on shifted /i/-word target trials by choosing a new set of shifted /i/-words that, when produced by the French talker, were not homophonous with existing unshifted /i/-words.

Experiment 4

In Experiment 4, we tested interpretation of shifted /i/-words that would not result in a pronunciation that was already associated with another English word when produced by the French talker. With this manipulation, we attempted to ensure that the most highly activated interpretation of the French talker's shifted /i/-words would be the correct interpretation.

We predicted that if listeners' performance in the previous experiments was negatively affected by interference from homophonous unshifted /i/-words, then in Experiment 4, we should see the pattern of results that we predicted (but did not obtain) for Experiments 1-3. If this was not the source of participants' difficulty in the previous experiments, then the results of Experiment 4 should be similar to those of Experiments 1-3.

Method

Participants—Sixty-five University of Illinois students participated in Experiment 4. Three additional participants were excluded due to technical errors. Participants received partial course credit for participating. All participants were native speakers of American English and had normal or corrected-to-normal vision and hearing. None of the participants reported having regular contact with anyone from Québec; none had participated in Experiments 1-3.

Stimuli—The stimuli were produced by the same female native French speaker from Experiments 1-3, but a different male native English speaker. The stimuli for the test phase consisted of seven sets of six English words. Each set of words contained an unshifted /i/-word (e.g., *wheeze*), a shifted /i/-word with the same onset (e.g., *weed*), and two pairs of filler words (see Appendix A for a full list of stimuli used in Experiment 4)². The shifted /i/-words ended in a variety of codas, as in the previous experiments, and the unshifted /i/-words ended in a voiced fricative (/z/ or /v/). Because, by design, the shifted /i/-words used in Experiment 4 had no unshifted /i/-word homophones (e.g., *wid* is not a word in English), it was not possible to directly compare the spectral properties of the French talker's vowels as we did in Experiments 1-3. However, given that the same French talker produced the stimuli used in all of the experiments, we expect that she continued to produce the vowels in the shifted /i/-words similarly for these stimuli. All visual stimuli were colorized clipart pictures selected in the same manner as in Experiments 1-3.

Equipment and Procedure—Like Experiments 1-3, Experiment 4 consisted of a training phase followed by a testing phase. The entire experiment lasted approximately 45 minutes.

Training: All participants completed a picture training session that used a procedure identical to that used in previous experiments. Each critical word and filler word was seen once during the training, for a total of 42 trials. All participants successfully learned the pictures to a 90% or better criterion on their first attempt.

In order to reduce the length of the experiment, exposure stories were not presented in Experiment 4. Instead, the testing phase itself was used as a vehicle for learning as, after all, the test situation provides ample opportunity for learning with feedback as the target word is always fully disambiguated by the coda consonant. In addition, other findings of robust accent learning have been reported in very similar paradigms that also did not use a pre-exposure story (Dahan, et al., 2008; Trude & Brown-Schmidt, 2012b).

²Several of our shifted stimuli were homophonous with very low-frequency English words (*keep-kip*, *peep-pip*, *seed-Sid*). Post-hoc analyses without these trials found that the pattern of effects was unchanged; therefore, we present the analyses containing the full data set.

Testing: The testing phase of Experiment 4 was the same as in Experiments 1-3. The testing phase of the experiment consisted of 224 randomly ordered trials (112 per talker). Each of the unshifted (e.g., *wheeze*) and shifted (e.g., *weed*) /i/-words was the target word four times per talker (8 times total), and each of the fillers was the target two times per talker (4 times total).

Results

The data from Experiment 4 were analyzed in the same manner as Experiments 1-3. Participants' accuracy was greater than 98% for all trial types.

Unshifted /i/-word target (e.g., *wheeze*), shifted /i/-word competitor (e.g., *weed*)

—Figure 8 presents the proportions of fixations to target, competitor, and filler words in the condition where the target was an unshifted /i/-word (e.g., *wheeze*) and the competitor was a shifted /i/-word (e.g., *weed*). Analysis on the time region from 0-200 ms revealed no baseline effects ($t = .63$). The ANOVA on the critical region of 200-1,600 ms showed a main effect of time, $F(14, 896) = 292.4, p < .001, (e = .207)$; $F(14, 84) = 23.8, p < .005, (e = .098)$, due to participants' increased likelihood to fixate the target as the trial progressed. The main effect of talker was marginal by subjects and items, $F(1, 64) = 3.7, p = .059$; $F(1, 6) = 3.9, p = .10$, with participants fixating the target slightly more on English talker than French talker trials ($M = 1.52$ and $M = 1.34$, respectively). A talker-by-time interaction, $F(14, 896) = 7.1, p < .001, (e = .284)$; $F(14, 84) = 3.7, p < .05, (e = .192)$, was due to a crossover pattern, in which participants initially fixated the target slightly more on French talker trials than English talker trials, then fixated the target more on English talker trials at later time points. Planned comparisons showed that the initial benefit on French talker trials was marginal by subjects at the 700 and 800 ms time windows ($ts > 1.89$); however, the later effect, in which participants fixated the target *less* with the French talker, was significant by subjects and items from 1,100-1,600 ms ($ts > 2.76$ and $ts > 2.46$, respectively), although marginal by items at the 1,300 ms window ($t = 1.86$).

Shifted /i/-word target (e.g., *weed*), unshifted /i/-word competitor (e.g., *wheeze*)

—Figure 9 presents the proportions of fixations to target, competitor, and filler words when the target was a shifted /i/-word (e.g., *weed*) and the competitor was an unshifted /i/-word (e.g., *wheeze*). A baseline analysis of the time region from 0-200 ms indicated that there was no baseline difference between talker conditions ($t = 1.43$). The ANOVA on the critical region revealed main effects of time, $F(14, 896) = 284.6, p < .001, (e = .249)$; $F(14, 84) = 82.5, p < .001, (e = .099)$ and talker, $F(1, 64) = 68.5, p < .001$; $F(1, 6) = 16.4, p < .01$, with participants fixating the target more when listening to the English talker than the French talker ($M = 1.54$ and $M = 1.31$, respectively). There was also a significant interaction between talker and time $F(14, 896) = 29.9, p < .001, (e = .323)$; $F(14, 84) = 13.9, p < .001, (e = .195)$. This interaction is characterized by a crossover effect, in which participants initially fixated the target more in the French talker condition, whereas later on, they fixated the target more in the English talker condition.

Planned comparisons revealed a significant talker effect by subject at the 500 ms window (French > English, $t = 2.35$), and a reversal from 800 to 1,600 ms by subject (English > French, $ts > 2.16$) and from 900 to 1,600 ms by items ($ts > 2.47$). Recall that we observed a similar pattern in Experiment 1; as in that experiment, this crossover is apparently due to participants' initial preference to fixate the competitor on English talker trials. This early competitor preference is likely a byproduct of the fact that the backward probability of having heard /i/ was greater for the unshifted /i/-words than the shifted /i/-words over the course of the experiment. One possibility for why the pattern only emerged in Experiments 1 and 4 is that the homogenous coda consonants of shifted /i/-words in Experiments 2 and 3

may have made participants less reliant on these backwards probabilities. However, unlike Experiment 1, participants in Experiment 4 do demonstrate a clear preference to fixate the target by the end of the French-talker trials, suggesting they were more confident in the ultimate interpretation of her speech.

Learning across the experiment—A learning analysis identical to those performed in Experiments 1-3 again revealed no learning effects on unshifted /i/-word (e.g., *wheeze*) trials, but a main effect of third, $F_1(1, 64) = 27.4, p < .001$; $F_2(1, 6) = 14.4, p < .01$, as well as a talker-by-third interaction for shifted /i/-word (e.g., *weed*) trials, $F_1(1, 64) = 7.2, p < .01$; $F_2(1, 6) = 28.4, p < .01$, indicating that participants improved more with the French talker than the English talker over the course of the experiment.

Discussion

The results of Experiment 4 provide tentative evidence that interference from homophonous unshifted /i/-words negatively affected participants' performance in Experiments 1-3. On trials with an unshifted /i/-word (e.g., *wheeze*) target, the data showed a brief numerical trend that was consistent with listeners applying knowledge of the French talker's accent in order to rule out a shifted /i/-word competitor, unlike the significant *deficit* in processing the French talker's speech observed in Experiments 1-3. However, even if the elimination of homophony in Experiment 4 allowed listeners to create a fuller representation of the French talker's pronunciation of the shifted words, the listeners' ability to use these representations was apparently quite limited. Strikingly, the numerical reduction in competition from shifted competitors observed in Experiment 4 was much smaller than that previously observed in studies of regionally-accented speech (Dahan et al., 2008; Trude & Brown-Schmidt, 2012a). One possible explanation for the increased difficulty in learning is the fact that the critical vowel shift was learned within the context of a foreign accent, which may present a more difficult learning situation due to increased variability in the speech signal associated with foreign accented speech (Nissen, Smith, Bradlow, & Bent, 2004; Wade, Jongman, & Sereno, 2007). We explore this possibility in Experiment 5.

Experiment 5

The goal of Experiment 5 was to test whether listeners' inability to learn a second-order constraint vowel shift in Experiments 1-4 was due to increased difficulty in processing foreign-accented speech. To test this hypothesis, we implemented a between-subjects design in which either the English talker or the French talker produced the critical vowel shift.

We predicted that if listeners' difficulty in learning the vowel shift in the previous experiments was due to increased variability in foreign-accented speech, then in Experiment 5, participants who hear the vowel shift produced by a native English talker should show more robust learning. If this was not the source of participants' difficulty in the previous experiments, then the results of Experiment 5 should be similar to those of Experiments 1-4.

Method

Participants—Sixty University of Illinois students participated in Experiment 5. Two additional participants were excluded due to technical errors. Participants received partial course credit for participating. All participants were native speakers of American English and had normal or corrected-to-normal vision and hearing. None of the participants reported having regular contact with anyone from Québec; none had participated in Experiments 1-4.

Stimuli—The stimuli were produced by the same female native French speaker from Experiments 1-4, but a different male native English speaker. For half of the participants, the

French talker produced the critical vowel shift just as in Experiments 1-4. The other half of the participants heard the English talker produce the shift, while the French talker produced all words with standard English vowel pronunciations. In order to elicit the correct pronunciations of the shifted /i/-words from the English talker, we provided him with orthographic representations of the shifted pronunciations (e.g., “wid” for *weed*). The words and visual stimuli were the same as those used in Experiment 4. As in Experiment 4, acoustic comparisons of the French talker’s vowels were not possible because, by design, the shifted /i/-word stimuli (e.g., *weed*) had no unshifted /i/-word homophones.

Equipment and Procedure—Like Experiments 1-4, Experiment 5 consisted of a training phase followed by a testing phase. The entire experiment lasted approximately 1 hour and 20 minutes.

Training: All participants completed a picture training session that used a procedure identical to that used in previous experiments. Each critical and filler word was seen once during the training, for a total of 42 trials. All participants successfully learned the pictures to a 90% or better criterion on their first attempt. As in Experiment 4, exposure stories were not presented.

Testing: The testing phase of Experiment 5 was the same as in previous experiments. The testing phase of the experiment consisted of 448 randomly ordered trials (224 per talker). Each of the unshifted /i/- (e.g., *wheeze*) and shifted /i/- (e.g., *weed*) words was the target word eight times per talker (16 times total), and each of the fillers was the target four times per talker (8 times total). The length of this experiment was doubled with respect to Experiment 4 in order to provide listeners with more exposure to the critical items and facilitate learning over the course of the experiment.

Results

The data from Experiment 5 were analyzed in the same way as Experiments 1-4, except that the identity of the talker producing the shift (English or French) was included as a factor. Participants’ accuracy was greater than 95% for all trial types.

Unshifted /i/-word target (e.g., *wheeze*), shifted /i/-word competitor (e.g., *weed*)

—Figures 10 and 11 present the proportions of fixations to target, competitor, and filler words when the target was an unshifted /i/-word (e.g., *wheeze*) and the competitor was a shifted /i/-word (e.g., *weed*) for participants who heard the vowel shift produced by, respectively, the French and the English talkers. Analyses on the time region from 0-200 ms revealed no baseline effects between talkers in either of the shift conditions (English talker vs. French talker producing shift) ($ts < 1.84$). The ANOVA on the critical region of 200-1,600 ms revealed main effects of time, $F_1(14, 812) = 436.1, p < .001, (e = .208)$; $F_2(14, 84) = 29.7, p < .001, (e = .127)$, as well as talker, $F_1(1, 58) = 6.7, p = .01$; $F_2(1, 6) = 10.2, p < .05$, with better performance on English talker than French talker trials, regardless of which talker produced the shift ($M = 1.64$ and $M = 1.47$, respectively). A talker-by-time interaction that was significant by subjects and marginal by items, $F_1(14, 812) = 4.5, p < .01, (e = .268)$; $F_2(14, 84) = 3.0, p = .10, (e = .124)$, was qualified by a talker-by-time-by-shift interaction, $F_1(14, 812) = 2.8, p < .05$; $F_2(14, 84) = 10.2, p = .01, (e = .192)$.

The nature of the 3-way interaction was such that, although performance on English talker trials was better than performance on French talker trials regardless of shift condition, performance on English talker trials was even better when the English talker was *not* the one producing the vowel shift. Conversely, performance on French talker trials was comparable across shift conditions. To explore this interaction, a series of planned 2×2 ANOVAs

compared talker and shift conditions at each 100 ms time window. The talker-by-shift interaction was marginal at 1,300 ms, significant from 1,400 to 1,500 ms, and marginal at 1,600 ms by subjects ($ps < .07$), and significant from 1,000 to 1,600 ms by items (marginal at 1,400 ms, all $ps < .06$). In all cases, in the French shift condition, participants performed better with the unshifted English talker than the shifted French talker. By contrast, in the English shift condition, performance was roughly equal with both talkers, or slightly better with the French talker in the later time windows.

Shifted /i/-word target (e.g., *weed*), unshifted /i/-word competitor (e.g., *wheeze*)

—Figures 12 and 13 present the proportions of fixations to target, competitor, and filler words in the condition where the target was a shifted /i/-word (e.g., *weed*) and the competitor was an unshifted /i/-word (e.g., *wheeze*) for participants who heard the vowel shift produced by, respectively, the French and the English talkers. Analyses of the time region from 0-200 ms indicated that there was no baseline difference between talkers in either shift condition ($ts < .54$). The ANOVA on the critical region revealed main effects of time $F(14, 812) = 415.3, p < .001, (\epsilon = .186)$; $F(14, 84) = 76.6, p < .001, (\epsilon = .116)$, and shift condition, $F(1, 58) = 7.3, p < .01$; $F(1, 6) = 23.6, p < .01$, with participants making more target fixations when listening to the unshifted talker. The interaction between talker and shift condition was found to be significant by both subjects and items, $F(1, 58) = 77.3, p < .001$; $F(1, 6) = 23.9, p < .01$. The interaction of talker and time was significant by subjects, $F(8, 512) = 11.3, p < .001, (\epsilon = .275)$, and the interaction of shift condition and time was significant by items, $F(14, 84) = 3.2, p < .05, (\epsilon = .216)$. These interactions were qualified by a significant three-way interaction of talker, time, and shift condition, $F(14, 812) = 42.7, p < .001$; $F(14, 84) = 22.8, p < .001, (\epsilon = .136)$.

The nature of this interaction was such that participants performed better when listening to the talker that did not produce the shift. Furthermore, comparing the target preference scores for a given talker across shift conditions reveals that performance decreased more with the shifted-English talker than the shifted-French talker (a difference of .62 for the English talker and .54 for the French talker). Planned 2×2 ANOVAs at each 100 ms time window revealed a significant talker-by-shift condition interaction from 500-600 ms by subjects only ($ps < .05$) in which participants performed about equally in the French shift condition but worse with the French talker in the English shift condition. This effect is driven by an initial preference to fixate the competitor when listening to the unshifted-French talker. This effect mirrors that seen in Experiment 1 and can be attributed to the same backward probability constraint. The talker-by-shift condition interaction was again significant from 800-1,600 ms ($ps < .001$ and .06, respectively). At these later time windows, participants perform better with whichever talker is *not* producing the shift, indicating difficulty in comprehending the shifted targets.

Learning across the experiment—Learning analyses similar to those for Experiments 1-4 were performed, and included shift-condition. The pattern of results mirrors that of the previous experiments. On the unshifted /i/-target (e.g., *wheeze*) trials, no significant learning effects were observed. However, on shifted /i/-target (e.g., *weed*) trials, we observe a main effect of third, $F(1, 58) = 15.5, p < .001$; $F(1, 6) = 10.9, p < .05$, as well as a two-way talker-by-third interaction significant by items, $F(1, 6) = 6.9, p < .05$. These effects were qualified by a three-way talker-by-third-by-shift interaction, in which participants showed greater improvement over the course of the study with the talker who produced the shift in their condition, $F(1, 58) = 11.8, p = .001$; $F(1, 6) = 10.8, p < .05$.

Discussion

The results of Experiment 5 showed that participants' difficulty in learning and using information about a second-order vowel shift was not entirely due to difficulty perceiving or learning about foreign-accented speech. On unshifted /i/-target (e.g., *wheeze*) trials, participants performed better with the English talker regardless of which talker was producing the vowel shift during the course of the experiment. This finding suggests that there may be something about foreign-accented speech that makes it more difficult for listeners to comprehend, perhaps due to its greater divergence from the speech that native English listeners are more used to hearing.

However, it should also be noted that on shifted /i/-target (e.g., *weed*) trials, listeners performed *worse* with the English talker when he was the one producing the shift. In fact, there appeared to be a greater detriment to performance when listening to the native English talker produce the shift, possibly due to surprise at hearing an otherwise native-sounding talker produce a very distinct and non-native-like vowel shift. This pattern is opposite of what is predicted in a case where the participants are able to learn about the critical vowel shift and use that information to rule out the shifted /i/-word (e.g., *weed*) competitor. Thus, it appears that listeners had difficulty learning and using information about the critical vowel shift even when it was produced by a native speaker. This indicates that participants' failure to learn the accent in Experiments 1-4 is not simply due to good-enough processing of foreign-accented speech, in which listeners simply loosen their expectations about foreign-accented speech due to their knowledge about the limited proficiency of the talker (Lev-Ari & Keysar, 2010), or an unwillingness or inability to learn a foreign-accented speaker's fine-grained phonetic details.

Participants' inability to adapt to the vowel shift even when it was produced by a native English talker suggests that adaptation to accents does not always occur quickly or easily, and that extended exposure to an accent may sometimes be needed in order for learning to occur (Witteman, Weber, & McQueen, 2013). These results contrast with those of Dahan et al. (2008) and Trude and Brown-Schmidt (2012a), who both found that listeners were able to learn and adapt to a similar second-order constraint in accented speech, as well as with research on adaptation to foreign accents that claims that adaptation occurs with minimal exposure (e.g., Clarke & Garrett, 2004). Why is it the case that we did not see such rapid adaptation in this study? It could be that having a foreign-accented speaker in the experiment at all, even when she is not producing the critical shift, makes adaptation more difficult due to the increased variability found in non-native speech.

Yet another possibility is that the participants in our studies had had less lifetime exposure to the critical accent than the participants in previous studies. For example, Clarke & Garrett (2004) tested native English speakers on Spanish- and Chinese-accented English, accents that a native English speaker would likely encounter with more frequency than the Québec French accent used here. Indeed, those authors note that many of their participants (University of Arizona undergraduates) had friends or family members who spoke with a Spanish accent. Likewise, Witteman, Weber, & McQueen (2013) found that only listeners who had previous experience with German-accented Dutch were initially primed by tokens produced with this accent. Similarly, participants in studies testing adaptation to regional dialects of American English may have been exposed to these dialects in everyday life, or via television programs or other media in their native language. In contrast, all of the participants in the current set of studies explicitly report being unfamiliar with the critical accent and having no contact with native Québec French speakers. The combination of significant improvements over the course of the experiment, particularly with the accented talker, with an inability to apply knowledge of the vowel-shift to rule out a lexical competitor suggests that learning may have been incomplete. A question for future research

is whether additional within-experiment exposure would be sufficient to show this effect, or whether more extensive real-world experience with an accent is necessary.

General Discussion

Many listeners report difficulty understanding foreign accented speech. Previous experiments have demonstrated that participants' understanding of foreign accented speech can improve with exposure, sometimes to near native-like levels (e.g., Bent & Bradlow, 2003; Bradlow & Bent, 2008; Clarke & Garrett, 2004; Witteman, Weber, & McQueen, 2013). However, most of the research in this area has focused on offline tasks, leaving open the possibility that while the ultimate interpretation of foreign-accented words may be correct, this *process* poses significant difficulty. Further, little existing research tests whether listeners can learn a particular feature of a non-native talker's accent, or which factors may help or hinder a listener in adapting to a foreign accent. In a series of experiments, we demonstrated that online interpretation of foreign accented speech is difficult for listeners, but that their difficulties can be mediated by different factors. The results of Experiments 2 and 3 suggest that the homogeneity of shifted words (as seen in previous studies on regional accent accommodation) does not appear to be the critical factor in helping listeners learn which words will be affected by the talker's accent. As shown in Experiment 4, the use of accent-shifted stimuli that have no homophone in standard English may help to slightly alleviate competition during online processing of accented speech, but that adaptation remains strikingly limited.

The results of all five experiments show that when an accent-shifted word itself is interpreted, listeners experience pronounced difficulty, possibly due to competing words from their own dialect of the language (e.g., Magnuson et al., 2007) or to a relative lack of exposure to the accent used in these studies. While the initial consonant-vowel combination present in the accented word uniquely identified the target and ruled out the competitor, we found that listeners were not able to take advantage of this unique identifying information, despite various simplifications of the stimulus set. We also found that this difficulty cannot be exclusively due to the increased variability or phonetic distance from standard English associated with a foreign accent, as indicated by listener's difficulty in identifying the accent-shifted words even when they were spoken by a native English talker. Lastly, perhaps the most important factor is experience: Across all five experiments, we observed significant improvement in interpreting these words over the course of the experiment, lending support to the possibility that listeners' poor performance was driven by a relative lack of exposure to the critical vowel shift.

Sub-phonemic vs. phonemic differences

An important consideration when interpreting the results of the present research with respect to previous findings of rapid adaptation to accents (e.g., Clarke & Garrett, 2004) is that we focused on a property of a foreign accent that, from the point of view of the listener, results in a specific *phonemic* change rather than a sub-phonemic difference. That is, the French talker's pronunciation of the /i/ vowel would be classified as an entirely different speech sound, /ɪ/, by most native English speakers. Because phonemes create contrastive differences in the language (e.g., the difference between *weak* and *wick*), learning to incorporate a sound that belongs to one category into an entirely different one may be difficult. While listeners are sensitive to sub-phonemic mismatches in the expected pronunciation of a given word (e.g., Dahan, Magnuson, Tanenhaus, & Hogan, 2001; Salverda, Dahan, & McQueen, 2003), we suspect that such sub-phonemic differences are likely to be less disruptive to a native speaker's speech processing because they will ultimately result in the correct mapping of the current sound to a phoneme category (although, see Witteman, Weber, &

McQueen, 2013, for a phonemic change in foreign-accented speech that was successfully learned).

If learning a phonemic change is difficult, why then were participants in previous research on adaptation to *regional* accents (Dahan et al, 2008; Trude & Brown-Schmidt, 2012a) able to learn a phonemic shift that was characteristic of a regional accent? One possibility is that participants had been exposed to these regional accents in every-day life prior to the experiment. By contrast, none of the participants had ever reported meeting a person from Quebec, and thus may have been less likely to have experienced that accent before. Indeed, the authors have informally noted several instances in which prominent American entertainment personalities exhibit features of these regional accents (e.g., Bob Dylan)³. The Quebec accent may have been sufficiently unfamiliar as to require more exposure prior to the listener being able to make useful and successful predictions based on learning of the second order constraint on the vowel shift. Consistent with this observation was the finding that despite a lack of a significant benefit from perceiving the accented talker, performance on accented-word trials did significantly improve across the course of the task *in all five experiments*. This robust learning effect shows that the accent of interest was clearly learnable within the constraints of the experimental situation, and that our measures were sensitive enough to pick up on this learning. Perhaps, then, with enough pre-exposure in day-to-day life, or with a much longer within-experiment exposure, participants would exhibit the robust learning and proactive use of the accent as previously observed with regional accents in a similar paradigm (Dahan et al, 2008; Trude & Brown-Schmidt, 2012a).

Relationship to previous research on adaptation to accents

Previous research on foreign-accented speech has largely focused on offline measures such as transcription. Indeed, listeners are found to improve at these tasks following exposure to a single talker or a group of different talkers with the same accent (e.g., Bradlow & Bent, 2008). However, in offline measures, it is difficult to gauge how quickly listeners are able to identify words in the accented talker's speech stream, as well as the processes that govern this ability. It is possible, for example, that when performing a transcription task (e.g., Bent & Bradlow, 2003; Bradlow & Bent, 2008; Sidaras, Alexander, & Nygaard, 2009), listeners initially misidentify words, but are able to revise their interpretation based on contextual information from the utterance or the delayed application of an explicit rule about the talker's accent. Because only the listener's final response is recorded, it is impossible to know how he or she arrived at that interpretation. Likewise, just because the participant produces a correct transcription does not mean that he or she did not have difficulty reaching that outcome.

Conversely, in the present experiments, the use of the visual-world eye-tracking paradigm allowed us to examine the time course of our participants' speech interpretation process. It is important to note that in all conditions of all five experiments, participants' accuracy in selecting the target picture was very high, so, as with previous research using offline measures, our participants eventually reached the correct interpretation of the French talker's speech. This aspect of the present results is highly consistent with previous findings on perception of foreign-accented speech (Bent & Bradlow, 2003; Bradlow & Bent, 2008; Clarke & Garrett, 2004; Sidaras, Alexander, & Nygaard, 2009): With little exposure, our participants readily clicked on the picture of *weak* when hearing [w₁ k] (Experiment 1), a picture of *wheat* when hearing [w₁ t] (Experiments 2-3), and a picture of a *weed* when hearing [w₁ d] (Experiments 4-5). However, the on-line analysis of listeners' eye-movements

³Specifically, the authors have noted that in the song "Like a Rolling Stone", Dylan sings of "Napoleon in [reɪ gʒ]", rather than "Napoleon in [rægz]".

reveal that this process was not an easy one, and that listeners experienced lingering uncertainty—evidenced by enduring fixations to both target and competitor—even when clicking on the correct object with above 95% accuracy. This finding shows that the *processing* of that speech may be slowed, and atypical *even when the ultimate interpretation is correct*.

Another difference between the present research and previous findings is that much of the previous research on adaptation to foreign-accented speech does not specify the particular features of the accent that required adaptation by listeners (e.g., Bradlow & Bent, 2008; Clarke & Garrett, 2004, Munro & Derwing, 1995). Thus, improved performance with exposure in these studies may have been due to adaptation to sub-phonemic aspects of the talker's accent. Conversely, our experiments were designed such that success was contingent upon participants learning a specific feature of the French talker's accent: the /i/ to [ɪ] shift before all consonants except voiced fricatives. Participants in our experiments must have been able to handle the auxiliary sub-phonemic differences produced by the French talker relatively easily, as demonstrated by their ability to identify the onset consonants of her speech successfully, but failed to readily adapt to the phonemic change, except (to a small extent) in the situation in which the phonemic change did not result in competition with an existing word in English (Experiments 4-5).

However, there is some evidence that listeners can process foreign-accented speech rapidly during on-line tasks. For example, Hanulíková and Weber (2012) found that Dutch and German second language (L2) speakers of English were more likely to fixate a *th*-initial word (e.g., *theft*) when the /θ/ was replaced by a sound more common for L2 speakers with the same language background as the participant (e.g., Dutch participants fixated [te ft] more than [se ft] because [t] is a more common pronunciation of /θ/ than [s] for native Dutch L2 English speakers). Likewise, Weber, Broersma, and Aoyagi (2011) found that in a cross-modal priming task, L2 English listeners were faster to make lexical decision judgments to printed words when their auditory primes were produced by another speaker with the same native language background (Dutch or Japanese), or if the accented pronunciation was confusable with a standard English pronunciation. Both of these studies show that knowledge about a foreign accent can indeed be applied in time to affect on-line processing of a single word; however, in both of these studies, a benefit was seen for an accent with which the listeners had previous experience or produced themselves. Conversely, in the present set of studies, little adaptation was seen when listeners were trained on a completely novel accent.

Our results nicely complement an eye-tracking study by Reinisch and Weber (2011), which examined Dutch listeners' learning of the non-canonical lexical stress patterns of a novel foreign accent. Their results showed that listeners who received evidence about the accented talker's stress patterns during training performed better at test than those who had received no evidence. However, performance on trials with non-canonically-stressed targets never exceeded performance on trials with canonically-stressed targets. These findings suggest that listeners were able to learn about the stress patterns used in a particular foreign accent; however, there was still a cost to processing the non-canonical word forms. Likewise, participants in our experiments demonstrated learning of the French talker's accent over the course of the experiment, but this learning was not enough to fully overcome the standard phonological representations of English that our participants were used to experiencing. The findings from these studies, together with our results, indicate that perhaps processing of accented speech can happen quickly and with little cost, but only after extensive exposure. Indeed, consistent with these findings of successful learning from familiar but foreign accents, analyses of learning across all five of our experiments revealed that listeners were capable of learning the present accent with sufficient exposure.

Conclusion

In conclusion, the results of the five experiments presented here demonstrate that listeners experience difficulty in learning and applying their knowledge of a vowel shift in foreign-accented speech during on-line speech processing, despite the fact that their ultimate interpretation was frequently correct. These results suggest that previous findings of successful interpretation of foreign-accented speech after minimal exposure (e.g., Bent & Bradlow, 2003; Bradlow & Bent, 2008; Clarke & Garrett, 2004; Sidaras, Alexander, & Nygaard, 2009; Weil, 2001) may mask more subtle difficulties revealed by the online method used here. We observed that learning improved somewhat when the critical vowel shift did not result in lexical competition, suggesting that aspects of an accent that result in phonemic changes and lexical conflict with the listener's dominant dialect may confer special challenges. We also observed that listeners performed poorly on shifted words regardless of the native or non-native status of the talker producing the shift, suggesting that adaptation to a specific vowel shift within the context of foreign-accented vs. native-accented speech does not present categorically different challenges for listeners. Finally, the fact that improvement on shifted words with increased exposure in all five experiments shows that the present accent was learnable, and further, strongly suggests that listeners' difficulties were primarily due to a need for extensive pre-exposure to a to-be-learned accent (e.g., Hanulíková & Weber, 2012; Weber, Broersma, & Aoyagi, 2011; Wittman, Weber, & McQueen, 2013).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors would like to thank Amélie Bernard, Greg Carmen, Siddarth Ponnala, and Jeet Raut for recording stimuli. This research was partially supported by NIH Grant T32-HD055272 and NSF-GRFP 2010084258 to AMT and NSF Grant 10-19161 to SBS.

References

- Bent T, Bradlow AR. The interlanguage speech intelligibility benefit. *Journal of the Acoustical Society of America*. 2003; 114:1600–1610. [PubMed: 14514213]
- Bradlow AR, Bent T. Perceptual adaptation to non-native speech. *Cognition*. 2008; 106:707–729. [PubMed: 17532315]
- Brainard DH. The Psychophysics Toolbox. *Spatial Vision*. 1997; 10:433–436. [PubMed: 9176952]
- Brown-Schmidt S. Beyond common and privileged: Gradient representations of common ground in real-time language use. *Language and Cognitive Processes*. 2012; 27:62–89.
- Clarke CM, Garrett MF. Rapid adaptation to foreign-accented English. *Journal of the Acoustical Society of America*. 2004; 116:3647–3658. [PubMed: 15658715]
- Dahan D, Drucker SJ, Scarborough RA. Talker adaptation in speech perception: Adjusting the signal or the representations? *Cognition*. 2008; 108:710–718. [PubMed: 18653175]
- Dahan D, Magnuson JS, Tanenhaus MK, Hogan EM. Subcategorical mismatches and the time course of lexical access: Evidence for lexical competition. *Language and Cognitive Processes*. 2001; 16:507–534.
- Evans BG, Iverson P. Vowel normalization for accent: An investigation of best exemplar locations in northern and southern British English sentences. *Journal of the Acoustical Society of America*. 2003; 115:352–361. [PubMed: 14759027]
- Ferguson HJ, Scheepers C, Sanford AJ. Expectations in counterfactual and theory of mind reasoning. *Language and Cognitive Processes*. 2010; 25:297–346.
- Flége JE. The production of “new” and “similar” phones in a foreign language: Evidence for the effect of equivalence classification. *Journal of Phonetics*. 1987; 15:47–65.

- Flege, JE. Second language learning: Theory, findings, and problems. In: Strange, W., editor. *Speech perception and linguistic experience: Issues in cross-language research*. Baltimore: York Press; 1995. p. 233-272.
- Flege JE, Frieda EM, Walley AC, Randazza LA. Lexical factors and segmental accuracy in second language speech production. *Studies in Second Language Acquisition*. 1998; 20:155–187.
- Flege JE, Schirru C, MacKay IRA. Interaction between the native and second language phonetic subsystems. *Speech Communication*. 2003; 40:467–491.
- Floccia C, Goslin J, Girard F, Konopczynski G. Does a regional accent perturb speech processing? *Journal of Experimental Psychology: Human Perception and Performance*. 2006; 32:1276–1293. [PubMed: 17002537]
- Gordon PC. Induction of rate-dependent processing by coarse-grained aspects of speech. *Perception & Psychophysics*. 1988; 43:137–146. [PubMed: 3340512]
- Hallett, PE. Eye movements. In: Kenneth, R Boff; Kaufman, Lloyd; Thomas, James P., editors. *Handbook of perception and human performance*. Vol. 1. Wiley; New York, New York: 1986. p. 10.1-10.112.
- Hanulíková A, Weber A. Sink positive: Linguistic experience with *th* substitutions influences nonnative word recognition. *Attention, Perception, & Psychophysics*. 2012; 74:613–629.
- Heller D, Grodner D, Tanenhaus MK. The role of perspective in identifying domains of reference. *Cognition*. 2008; 108:831–836. [PubMed: 18586232]
- Kraljic T, Samuel AG. Perceptual adjustments to multiple speakers. *Journal of Memory and Language*. 2007; 56:1–15.
- Lev-Ari, S.; Keysar, B. When chocolate cakes are “good-enough” pies: Adjusting manner of processing to non-native speakers; Poster presented at the 23rd Annual CUNY Conference on Human Sentence Processing; New York, NY. 2010 Mar.
- Lieberman AM, Cooper FS, Shankweiler DP, Studdert-Kennedy M. Perception of the speech code. *Psychological Review*. 1967; 74:431–461. [PubMed: 4170865]
- Magnuson JS, Dixon JA, Tanenhaus MK, Aslin RN. The dynamics of lexical competition during spoken word recognition. *Cognitive Science*. 2007; 31:133–156. [PubMed: 21635290]
- Maye J, Aslin RN, Tanenhaus MK. The Weckud Wetch of the West: Lexical adaptation to a novel accent. *Cognitive Science*. 2008; 32:543–562. [PubMed: 21635345]
- Munro MJ, Derwing TM. Processing time, accent, and comprehensibility in the perception of native and foreign-accented speech. *Language and Speech*. 1995; 38:289–306. [PubMed: 8816082]
- Nissen SL, Smith BL, Bradlow A, Bent T. Accuracy and variability in vowel targets produced by native and non-native speakers of English. *Journal of the Acoustical Society of America*. 2004; 116:2604.
- Pelli DG. The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*. 1997; 10:437–442. [PubMed: 9176953]
- Peterson GE, Barney HL. Control methods used in a study of vowels. *Journal of the Acoustical Society of America*. 1952; 24:175–184.
- Reinisch, E.; Weber, A. Adapting to lexical stress in a foreign accent. In: Lee, W-S.; Zee, E., editors. *Proceedings of the 17th International Congress of Phonetic Sciences*; 2011. p. 1678-1681.
- Salverda AP, Dahan D, McQueen JM. The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension. *Cognition*. 2003; 90:51–89. [PubMed: 14597270]
- Samuel A. Phoneme Restoration. *Language and Cognitive Processes*. 1996; 11:647–653.
- Sidaras SK, Alexander JED, Nygaard LC. Perceptual learning of systematic variation in Spanish-accented speech. *Journal of the Acoustical Society of America*. 2009; 125:3306–3316. [PubMed: 19425672]
- Sumner M, Samuel AG. The effect of experience on the perception and representation of dialect variants. *Journal of Memory and Language*. 2009; 60:487–501.
- Trude AM, Brown-Schmidt S. Talker-specific perceptual adaptation during online speech perception. *Language and Cognitive Processes*. 2012a; 27:979–1001.

- Trude, AM.; Brown-Schmidt, S. Can Listeners Use Creaky Voice to Constrain Lexical Interpretation?; Poster presented at the 18th Annual Conference on Architectures and Mechanisms for Language Processing; Riva Del Garda, Italy. 2012b.
- Wade T, Jongman A, Sereno J. Effects of acoustic variability in the perceptual learning of non-native-accented speech sounds. *Phonetica*. 2007; 64:122–144. [PubMed: 17914280]
- Weber A, Broersma M, Aoyagi M. Spoken-word recognition in foreign-accented speech by L2 listeners. *Journal of Phonetics*. 2011; 39:479–491.
- Weil, SA. Foreign accented speech: Adaptation and generalization (Master's thesis). The Ohio State University; Columbus, OH: 2001.
- Witteman MJ, Weber A, McQueen JM. Foreign accent strength and listener familiarity with an accent codetermine speed of perceptual adaptation. *Attention, Perception, & Psychophysics*. 2013; 75:537–556.

- Listeners' online adaptation to foreign-accented speech was tested with eye-tracking.
- Despite high accuracy, listeners' eye movements revealed processing difficulties.
- Performance was best when accented words did not have a standard English homophone.
- Offline tasks showing fast accent adaptation mask difficulties in online processing.

Figure 1.

Example of a display on a critical trial. The unshifted /i/-word is *wheeze* (upper left), and the shifted /i/-word is *weak* (upper right). On critical trials, participants heard either “Select wheeze” or “Select weak”.

Figure 2.

Proportion of fixations to target, competitor, and fillers on English- and French-talker trials with an unshifted /i/-word target (e.g., *wheeze*) and shifted /i/-word competitor (e.g., *weak*) over time in Experiment 1.

Figure 3.

Proportion of fixations to target, competitor, and fillers on English- and French-talker trials with a shifted /i/-word target (e.g., *weak*) and unshifted /i/-word competitor (e.g., *wheeze*) over time in Experiment 1.

Figure 4.

Proportion of fixations to target, competitor, and fillers on English- and French-talker trials with an unshifted /i/-word target (e.g., *wheeze*) and shifted /i/-word competitor (e.g., *wheat*) as a function of time in Experiment 2.

Figure 5.

Proportion of fixations to target, competitor, and fillers on English- and French-talker trials with a shifted /i/-word target (e.g., *wheat*) and unshifted /i/-word competitor (e.g., *wheeze*) as a function of time in Experiment 2.

Figure 6.

Proportion of fixations to target, competitor, and fillers on English- and French-talker trials with an unshifted /i/-word target (e.g., *wheeze*) and shifted /i/-word competitor (e.g., *wheat*) as a function of time in Experiment 3.

Figure 7.

Proportion of fixations to target, competitor, and fillers on English- and French-talker trials with a shifted /i/-word target (e.g., *wheat*) and unshifted /i/-word competitor (e.g., *wheeze*) as a function of time in Experiment 3.

Figure 8.

Proportion of fixations to target, competitor, and fillers on English- and French-talker trials with an unshifted /i/-word target (e.g., *wheeze*) and shifted /i/-word competitor (e.g., *weed*) as a function of time in Experiment 4.

Figure 9.

Proportion of fixations to target, competitor, and fillers on English- and French-talker trials with a shifted /i/-word target (e.g., *weed*) and unshifted /i/-word competitor (e.g., *wheeze*) as a function of time in Experiment 4.

Figure 10.

Proportion of fixations to target, competitor, and fillers on English- and French-talker trials with an unshifted /i/-word target (e.g., *wheeze*) and shifted /i/-word competitor (e.g., *weed*) in the French-shift condition as a function of time in Experiment 5.

Figure 11.

Proportion of fixations to target, competitor, and fillers on English- and French-talker trials with an unshifted /i/-word target (e.g., *wheeze*) and shifted /i/-word competitor (e.g., *weed*) in the English-shift condition as a function of time in Experiment 5.

Figure 12.

Proportion of fixations to target, competitor, and fillers on English- and French-talker trials with a shifted /i/-word target (e.g., *weed*) and unshifted /i/-word competitor (e.g., *wheeze*) in the French-shift condition as a function of time in Experiment 5.

Figure 13.

Proportion of fixations to target, competitor, and fillers on English- and French-talker trials with a shifted /i/-word target (e.g., *weed*) and unshifted /i/-word competitor (e.g., *wheeze*) in the English-shift condition as a function of time in Experiment 5.

Acoustic measurements for French and English talkers' productions of unshifted /i/-words, shifted /i/-words, and unshifted /u/-fillers used in Expt. 1 (data shown are means and 95% confidence intervals of the mean, in brackets).

Table 1

Acoustic Measurement	Unshifted /i/-words (e.g., <i>wheeze</i>)		Shifted /i/-words (e.g., <i>weak</i>)		Unshifted /u/-fillers (e.g., <i>wick</i>)	
	French	English	French	English	French	English
F1 (Hz)	339.5 [310.9, 368.1]	274.0 [265.8, 282.2]	646.7 [588.8, 704.6]	301.2 [282.7, 319.7]	642.2 [584.4, 700.0]	483.8 [470.5, 497.1]
F2 (Hz)	2725.6 [2423.9, 3027.1]	2237.8 [2210.7, 2264.9]	2215.3 [2147.9, 2282.7]	2266.6 [2208.1, 2325.1]	2240.0 [2154.2, 2325.8]	1676.8 [1605.1, 1748.5]
Vowel Duration (ms)	447.5 [393.6, 501.3]	441.2 [412.8, 469.6]	181.9 [153.8, 210.0]	187.3 [161.1, 213.6]	193.4 [163.5, 223.3]	173.5 [150.7, 196.3]

Table 2

Critical trial types. Column 1 indicates the target and competitor word types. Column 2 indicates the talker identity. Columns 3-4 indicate how the talker pronounces the target and competitor words. Column 5 indicates whether the talker produces the target and competitor word with the same vowel. A complementary set of filler trials included *wick* (rather than *weak*) as either target or competitor word.

Trial type	Talker	Target word	Competitor picture	Same vowel?
Unshifted /i/ target; Shifted /i/ competitor	English	<i>wheeze</i> [wiz]	<i>weak</i> [wik]	Yes
	French	<i>wheeze</i> [wiz]	<i>weak</i> [wɪ k]	No
Shifted /i/ target; Unshifted /i/ competitor	English	<i>weak</i> [wik]	<i>wheeze</i> [wiz]	Yes
	French	<i>weak</i> [wɪ k]	<i>wheeze</i> [wiz]	No

Acoustic measurements for French and English talkers' productions of unshifted /i/-words, shifted /i/-words, and unshifted /ɪ/-fillers used in Expt. 2 (data shown are means and 95% confidence intervals of the mean, in brackets).

Table 3

Acoustic Measurement	Unshifted /i/-words (e.g., <i>wheeze</i>)		Shifted /i/-words (e.g., <i>heat</i>)		Unshifted /ɪ/-fillers (e.g., <i>wit</i>)	
	French	English	French	English	French	English
F1 (Hz)	357.4 [326.0, 388.9]	274.9 [264.0, 285.7]	648.9 [604.8, 692.9]	296.3 [287.6, 305.0]	653.3 [597.9, 708.7]	490.1 [461.4, 518.9]
F2 (Hz)	2654.1 [2208.5, 3099.8]	2226.9 [2192.9, 2260.8]	2109.9 [1856.5, 2363.2]	2282.3 [2233.6, 2331.0]	2202.6 [2127.6, 2277.5]	1798.0 [1739.7, 1856.3]
Vowel Duration (ms)	436.0 [379.1, 492.9]	423.6 [399.1, 448.1]	190.9 [156.0, 225.8]	177.0 [166.4, 187.7]	198.2 [158.5, 238.0]	168.8 [150.3, 187.4]