Research Article

Training-induced pattern-specific phonetic adjustments by first and second language listeners

Angela Cooper *, Ann Bradlow

Department of Linguistics, Northwestern University, 2016 Sheridan Rd., Evanston, IL 60208, USA

ARTICLE INFO

Article history:
Received 24 March 2017
Received in revised form 4 January 2018
Accepted 20 February 2018

Keywords:
Perceptual learning
Bilingualism
Spoken word recognition
Dutch
English
Foreign accents

ABSTRACT

The current study investigated the phonetic adjustment mechanisms that underlie perceptual adaptation in first and second language (Dutch-English) listeners by exposing them to a novel English accent containing controlled deviations from the standard accent (e.g. /i/-to-/ɪ/ yielding /krɪm/ instead of /krim/ for 'cream'). These deviations involved contrasts that either were contrastive or were not contrastive in Dutch. Following accent exposure with disambiguating feedback, listeners completed lexical decision and word identification tasks. Both native and second language listeners demonstrated adaptation, evidenced by higher lexical endorsement rates and word identification accuracy than untrained control listeners for items containing trained accent patterns. However, for L2 listeners, adaptation was modulated by the phonemic contrast, that is, whether or not it was contrastive in the listeners' native language. Specifically, the training-induced criterion loosening for the L2 listeners was limited to contrasts that exist in both their L1, Dutch, and L2, English. For contrasts that are either absent or neutralized in Dutch, the L2 listeners demonstrated relatively loose pre-training criteria compared to L1 listeners. The results indicate that accent exposure induces both a general increase in tolerance for atypical speech input as well as targeted adjustments to specific categories for both L1 and L2 listeners.

© 2018 Elsevier Ltd. All rights reserved.

1. Introduction

1.1. Perceptual adaptation for native listeners

Native (L1) listeners of a language possess a remarkably flexible perceptual system that enables them to extract linguistic information from degraded or impoverished speech signals, to identify different talkers with a high degree of accuracy, and to adapt to variability that arises as a function of differences in talker or accent (Cutler, 2012). Adaptation to talker-specific characteristics, such as a foreign accent, occurs rapidly, within a few minutes of exposure (Clarke & Garrett, 2004) and can subsist for several days without any intervening exposure (Eisner & McQueen, 2006; Kraljic & Samuel, 2005). Prior research has suggested that adaptation to patterns of deviation from native-accented norms arises from listeners' ability to utilize contextual information, including lexical (e.g., Eisner & McQueen, 2005; Kraljic & Samuel, 2007; Zhang & Samuel, 2014), phonotactic (Cutler, McQueen, Butterfield, Norris, & Planck, 2008) and visual information (e.g., Bertelson, Vroomen, & De Gelder, 2003) to interpret these deviations (e.g., determine that [wɪʃ] is likely witch /wɪʃ/] and adjust the relevant phoneme category boundaries as necessary. Moreover, foreign-accented speech also appears to be processed gradiently, whereby the strength of the accent influences the speed and accuracy of word recognition (Porretta, Tucker, & Järvikiv, 2016).

The specificity of perceptual learning has also been the subject of investigation (e.g., Baese-Berk, Bradlow, & Wright, 2013; Bradlow & Bent, 2008; Eisner & McQueen, 2005; Kraljic & Samuel, 2005, 2006; Kraljic & Samuel, 2007; Reinisch & Holt, 2014; Reinisch & Mitterer, 2016; Sidaras, Alexander, & Nygaard, 2009), with a number of factors implicated in whether or not listeners will generalize learning to novel talkers, including the type of phonetic contrast involved, the amount of relevant variation in the signal and the acoustic similarity of the talkers. For example, when presented with multiple talkers who share the same foreign accent, perceptual learning was more likely to generalize to a novel talker of the same accent than if exposed to a single talker (Bradlow & Bent, 2008; Sidaras et al., 2009). High variability training is
posed to promote generalizable learning, as it allows for the extraction of systematic patterns of deviations from native-accented norms shared by the talkers, resulting in more robust and generalizable adjustments. However, there have been cases where single talker training on a regional accent characteristic has been found to transfer to a novel talker, which has been attributed to the acoustic similarity between the trained and novel talkers (e.g., Reinsch & Holt, 2014).

Importantly, prior work has demonstrated systematic (or constrained) adaptation as opposed to unconstrained across-the-board criterion loosening. Maye et al. (2008) examined native listeners’ adaptation to a novel accent of English by comparing performance in an auditory lexical decision task following two different exposure phases: (1) a 20-min story produced with a Standard American English accent, and (2) the same story where all instances of front vowel pronunciations were lowered. Following exposure to the novel accent, a significant increase in lexical endorsement rates in the lexical decision task was found for lowered front-vowel items but not raised front-vowel items. This suggests that adaptation processes involved direction-specific, targeted phonetic adjustments rather than relatively unconstrained, general category broadening. Weatherholtz (2015), using a similar paradigm, found that the enhanced word recognition following accent exposure generalized to novel items as well as to novel talkers, despite listeners being initially exposed to only a single talker. Unlike Maye et al. (2008), listeners were also found to generalize their exposure to untrained, structurally-related vowel chain shifts. That is, following exposure to a system of back vowel lowering, enhanced word recognition was found when tested on a system of back vowel raising or front vowel lowering. However, exposure to a system of back vowel raising did not yield generalization to back vowel lowered pronunciations, suggesting that adaptation processes may involve a combination of general category broadening and targeted category shifts. Evidence for this mechanism of general category expansion has also been implicated in accent adaptation with young children (Schmale, Cristia, & Seidl, 2012; Schmale, Seidl, & Cristià, 2015).

1.2. Non-native speech perception and learning

Despite the precision and flexibility of native language listening, speech perception in one’s second language can be a challenging task, arising from difficulties at multiple levels of linguistic processing (e.g., Best & Tyler, 2007; Broersma & Cutler, 2008; Flege, 1995). As a result of inaccurate L2 phoneme perception, L2 spoken word recognition can become problematic. Words such as peck and pack for Dutch listeners or rice and lice for Japanese listeners are often indistinguishable from each other (Broersma & Cutler, 2008; Logan, Lively, & Pisoni, 1991). L2 listeners have the added challenge of contending with more lexical competitors during word recognition than L1 listeners, as a product of the activation of “phantom” competitors (Broersma & Cutler, 2008; Cutler & Broersma, 2005). For instance, in a lexical decision task, near-word items such as “flide” or “shib” would be considered non-words by native English listeners but more often as real words by native Dutch listeners (perceiving them as flight and ship, respectively), as a result of voicing not being distinctive word-finally in Dutch. Indeed, in a cross-modal priming paradigm, Dutch listeners’ perception of the targets were significantly facilitated both when the prime item matched (flight–FLIGHT) as well as when the prime item was a near-word (flide–FLIGHT), whereas this facilitation was only found in the matched condition for English listeners. A similar pattern of results was found with the /æi/-/ɛi/ contrast, which is also not distinctive in Dutch (Cutler & Broersma, 2005). This indicates that perceptual phonetic confusions and phantom activation can lead to an increase in the amount of lexical competition with which an L2 listener has to contend, and a greater degree of lexical competition has been demonstrated to yield slower word recognition (e.g., Norris, McQueen, & Cutler, 1995).

However, there is evidence to suggest that L2 listeners can achieve comparable performance (or even surpass) native listeners when listening to foreign-accented speech (that is, speech produced by other L2 speakers), particularly when the talker and listener share an L1 background (Bent & Bradlow, 2003; Hayes-Harb, Smith, Bent, & Bradlow, 2008; Imai, Walley, & Flege, 2005; van Wijngaarden, 2001; Xie & Fowler, 2013). For example, using a cross-modal priming paradigm, Weber, Broersma, and Aoyagi (2011) found facilitated processing when the acoustic manifestation of the foreign-accented prime item aligned with the accent of the listeners (Japanese vs. Dutch). For example, the item /ski/ primed the English word act for Dutch listeners, whereas /ˈaktoʊ/ primed act for Japanese listeners (and did not for Dutch listeners). Similarly, Hanulíková and Weber (2012) also reported the influence of linguistic experience on the perception of foreign-accented pronunciation variants. The eye movements of German and Dutch learners of English were tracked when perceiving words containing three variants (/is/, /fi/, /it/) of the pronunciation of the English interdental fricative /θ/. Despite /θ/ being the most perceptually confusable with /t/, listeners’ looking preferences aligned with the pronunciation variant most frequently produced by listeners of the different language groups (German-accented English is characterized by /s/-substitutions, while Dutch-accented English is marked by /t/-substitutions). Knowledge or familiarity with the particular deviation patterns that result from a specific L1–L2 language pair may better equip listeners to interpret the speech produced by a speaker with this particular language background.

As noted above, one of the hallmarks of L1 listening is the ability to flexibly adapt to variable speech input (Cutler, 2012), and within the context of the L2 speech perception research reviewed above, recent work has begun to examine whether and how perceptual adaptation works for L2 listeners (e.g., Mitterer & McQueen, 2009; Reinisch, Weber, & Mitterer, 2012; Schertz, Cho, Lotto, & Warner, 2015; Weber, Di Betta, & McQueen, 2014). Mitterer and McQueen (2009) reported that Dutch listeners exposed to Scottish or Australian English accented speech on television were able to leverage English language subtleties to improve their comprehension relative to control listeners who either received Dutch subtitles or no subtitles. Listeners who received English subtitles were significantly more accurate at repeating back phrases produced by Australian- and Scottish-accented speakers as compared to the control groups who did not receive any training. As these L2 listeners demonstrated an ability to utilize lexical (as well as audio-visual and other sub- and supra-lexical) cues to adapt
to the unfamiliar English accent, this suggests that adaptation in a second language may draw upon similar resources as adaptation in the native language.

Of relevance to the present work, Grohe and Weber (2016) trained English monolinguals and German-English bilinguals on an artificial English accent produced by a German speaker. In this accent, the interdental fricative /θ/ was replaced with the voiceless stop /t/, which is an uncommon substitution in German-accented English. Participants were either exposed to a story with this accent or had to produce the story with this accent (produce t-substitutions). Auditory lexical decision results revealed that the L2 bilinguals showed more robust adaptation relative to the L1 listeners, with production training proving to be more facilitative than perception training for the L2 listeners. The authors posit that L2 listeners possess more flexible representations as a result of having higher production variability.

1.3. The current study

The present work sought to understand the phonetic adjustment mechanisms that underlie perceptual adaptation in first and second language listeners. When encountering an unfamiliar accent pattern, does adaptation involve making a targeted, context-specific shift (e.g., Maye et al., 2008), or does it involve a more system-wide general relaxing of criteria for what constitutes an acceptable match between acoustic input and stored representation (e.g., Schmale et al., 2015), or a combination of these mechanisms?

Moreover, prior research on exposure-induced phonetic adjustments with L2 listeners has predominantly focused either on a single contrast that exists in both L1 and L2 languages (e.g., Reinisch et al., 2012) or not controlled for the segmental content of the regionally-accented speech (e.g., Mitterer & McQueen, 2009). Natural speech produced by a regionally- or foreign-accented speaker will typically contain a range of different deviations from standard-accented norms, potentially involving a variety of phonemic categories that may or may not be well-established for an L2 listener. The question remains whether and how L2 listeners accommodate pronunciation variation for contrasts that both exist and do not exist in their first language. How L2 listeners adapt to specific phonetic deviations of unfamiliar accented speech in their second language may provide insight into the circumstances under which the different phonetic adjustment mechanisms may be employed and the balance between them.

To examine this issue, L1 English and L2 Dutch-English listeners completed a two-phase experiment that directly tested training-induced phonetic adjustments to deviant word productions. Specifically, this paradigm was comprised of passive exposure to a novel English accent, where listeners were provided with a visual linguistic context (e.g., the word west printed on the screen) and heard an accented production of a deviant segment in a word context (e.g., /waest/). Following exposure, they completed lexical decision and word identification tasks. The lexical decision task enabled us to examine whether exposure to a novel accent would shift listeners’ willingness to endorse accented items as being real English words. Furthermore, it allowed us to examine whether it reflected accent pattern-specific adjustments, manifesting as an increase in lexical endorsement rates to only items containing the trained accent patterns, or a more general relaxing of permissible matches between speech input and lexical representations, manifesting as an increase in endorsement rates to both trained pattern items and non-words. The word identification task provided insight into not only whether or not listeners would consider an accented item to be a word but also whether they could accurately identify it based on the specific pattern of deviation that they experienced during exposure.

By virtue of having less exposure to the target language, L2 listeners may hold a relatively higher degree of linguistic uncertainty—a general baseline uncertainty about the language at all levels but also specific uncertainty about the distributions of phonetic cues to particular phonemic contrasts not present in their L1. This general linguistic uncertainty about the language could constrain adaptation (Kleinschmidt & Jaeger, 2015), whereby listeners may require a greater amount of evidence before being willing to make perceptual adjustments. If second language listening involves a higher degree of general linguistic uncertainty, thereby inhibiting adaptation, this would predict an overall smaller increase in lexical endorsement rates for L2 relative to L1 listeners. Additionally, their specific uncertainty about particular L2 phonemic contrasts would be predicted to have a significant impact on adaptation. If items containing contrasts that are neutralized in the L1 provoke higher uncertainty for L2 listeners, then we would predict that the L2 listeners would be slower to adapt to items containing contrasts that are phonological in English but not meaningful in Dutch (i.e., English only contrasts) relative to items containing contrasts that exist in both English and Dutch (i.e., Dutch/English contrasts).

Alternatively, since L2 listeners’ experience with speech in their L2 likely involved exposure to quite variable pronunciations (e.g., the balance between L1- and L2-accented pronunciations is likely to be more even for L2 listeners than for L1 listeners), L2 listeners may be generally more flexible in their mappings of variable speech input onto stored representations (Weber et al., 2014). This may yield an overall larger increase in the rate of lexical endorsements for non-standard pronunciations during training in the L2 for second versus first language listeners. Moreover, their familiarity with hearing variable pronunciations of English only contrasts, as a product of experience hearing Dutch-accented English, may result in comparable or even greater adaptation to English only items relative to Dutch/English items. Indeed, in the present study, some of the items containing English only items were produced in the same ways as some Dutch-accented English speakers would produce them. Therefore, their prior experience with these specific cue distributions could be leveraged to more efficiently update their beliefs about the distributions of the Non-Standard American English (NSAE) accent in this study.

2. Methods

2.1. Overview of experimental setup

The current study exposed L1 English and L2 Dutch-English listeners to a novel accent of English with items containing phonemic contrasts that exist in Dutch as well as items
that contain contrasts that do not exist in Dutch or are neutralized in certain Dutch contexts. This NSAE (Non-Standard American English) accent was constructed to contain both vowel deviation patterns (a word such as *bleak* pronounced as [blik]) as well as consonant patterns (*throw* pronounced as [tou]). Exposure consisted of a written presentation of either a word, matching the target item, or a semantic context sentence, predictive of the target, followed by an auditory presentation of the target item. Target items were produced by either a single talker or divided between multiple talkers throughout training. Following exposure, listeners were tested with lexical decision and word identification tasks and compared with listeners who had not received NSAE-accented exposure. L2 listeners also completed an additional phonetic assessment word identification task, which enabled us to determine whether listeners were capable of accurately identifying items containing the phonemes involved in the accent to which they were exposed during training. Fig. 1 provides an overview of the experimental setup. For the lexical decision task, adaptation would manifest as a higher proportion of NSAE-accented items (considered non-words in a Standard American English accent) endorsed as lexical items by trained listeners relative to control listeners. Additionally, for the word identification task, a higher proportion of NSAE-accented items transcribed based on the novel accent patterns (e.g., [blik] identified as *bleak*) would be indicative of adaptation.

2.2. Participants

One hundred and fifteen native American English listeners, which included 35 participants (*F* = 23; *M* age = 19.3 years) tested in the lab and 80 participants (*F* = 50; *M* age = 36.6 years) tested on Amazon Mechanical Turk (an online service that provides an on-demand human workforce for a variety of different tasks), received course credit or were paid for their participation (Table 1). American English listeners were defined as having English as their primary language prior to school (i.e., prior to approximately 6 years of age) and as the primary language of instruction during school. None had any prior experience with Dutch.

Ninety-four L1 Dutch-L2 English listeners, including 41 participants (*F* = 30; *M* age = 20.4 years) tested in the lab and 53 participants (*F* = 45; *M* age = 21.3 years) tested online, were included in the study and were paid 8 Euros for their participation. Both lab and online participants were native speakers of Dutch studying at Radboud University Nijmegen. Online participants were recruited via email and provided a link to the identical online experiment platform as the L1 subjects and the in-lab Dutch listeners. To increase the potential differences between L1 and L2 groups, we recruited Dutch listeners who were not pursuing a Master’s in English or following their Bachelor’s or Master’s degrees in English, such that their dominant language of instruction at the university was not English. Indeed, Dutch was reported as the language of instruction during their primary, secondary and post-secondary education. Participants reported acquiring English after the age of 7 and had been learning English for on average 7 years. Their average self-reported proficiency (on a scale of 1–10, 10 being native) in Dutch was 9.3 and in English was 6.7. The majority of the participants also had experience learning a third language, including German (*n* = 46), French (*n* = 26), Spanish (*n* = 3), and Greek (*n* = 1). In-lab participants were tested in a sound-attenuated booth either at Northwestern University or the Max Planck Institute for Psycholinguistics. All participants self-reported no hearing impairments at the time of testing.

2.3. Procedure

Participants in the training conditions completed two blocks of training, each preceded by a probe lexical decision task. Following training, they completed two test tasks: (1) lexical decision, and (2) word identification. Consistent with recent work (Grohe & Weber, 2016), participants in a control condition only completed the test tasks (lexical decision and word identification). L2 listeners also received a phonetic assessment task after they completed these test tasks. Trained listeners took approximately 45–60 min to complete these tasks. All participants were informed that they would be listening to someone with an accent (trained listeners were further instructed that they would be receiving training to improve their ability to understand the accent).

Training consisted of 2 phases of passive exposure to a constructed (as opposed to naturally-occurring) novel accent (Non-Standard American English, NSAE) accompanied by one of two types of feedback. Participants would first view a written presentation of either the target word (e.g., *dress*) or a context sentence (e.g., *She tried on a ___*). Once they had finished reading the display, they would click “NEXT”, at which point the written context would disappear and they would hear an auditory presentation of the target word produced in the NSAE accent. This paradigm is similar to ones used in research on adaptation to noise-vocoded speech, where listeners would be presented with a distorted word or phrase followed by a clear, undistorted presentation of that stimulus (e.g., Hervais-Adelman, Davis, Johnsrude, & Carlyon, 2008). Disambiguating feedback information was visually-presented to ensure that participants who only received target word feedback would not have less exposure to the accent relative to those who received context sentences (which naturally have more material). All items were produced either by a single talker or were divided between 4 different talkers (blocked by talker).

Participants completed two probe lexical decision blocks (each containing 36 trials): one immediately prior to training and one after the first phase of training. Each item, none of which were heard during training, was presented individually. Participants were asked to respond as quickly as possible as to whether they thought the item was a word or a nonword of English. Participants used a mouse to click one of two radial response options on the screen. They were instructed that the speaker in these probe blocks was a talker who spoke with the same accent they heard during training. Item presentation was randomized within each block, and the order of the probe blocks was counterbalanced across participants.

Following the training and lexical decision probes, participants proceeded to the test phase, which consisted of two separate tests, lexical decision and word identification. In each test, the stimulus items were spoken by two talkers, a trained talker (i.e., a talker from the training phase) and a generalization talker (i.e., a novel talker). Participants were instructed that
the talker was either someone they had heard during training or a novel talker. They were also informed that the novel talker spoke with the same accent as they had been exposed to during training. The lexical decision task in this test phase followed exactly the same procedure as the probe task, in which participants responded "word" or "non-word" to each individually presented item. A total of 189 trials were presented, including trained as well as novel items, and was blocked by talker, with the trained talker always preceding the generalization talker. Following the lexical decision test, participants completed a word identification task, which consisted of a total of 132 trials of novel items. Similar to the lexical decision task, trials were blocked by talker (66 trials each), and the trained talker block was presented before the generalization talker block. Each item was presented individually, and participants were asked to type the word they heard. If they did not believe the word to be a real word, they could type 'X'. There was no limit on response time.

Finally, the phonetic assessment task consisted of a total of 30 randomized trials and was completed at the end of the experiment. This task was administered in order to establish L2 listeners’ baseline ability to identify words containing the segments affected by the NSAE accent (e.g., /æ/, /æ/, /æ/)—that is, to determine whether they might identify such words as "bed" and "thought" as "bad" and "taught", respectively. This task enabled us to examine the extent to which L2 listeners had established L2 categories, particularly for difficult contrasts that are not present in Dutch. Each item was presented to listeners individually, and they were asked to transcribe the item they heard. Listeners were informed that the talker would be speaking with a Standard American English accent, and that all presented items were real words. There was no limit on response time.

Table 1
Participant breakdown by condition and language background.

<table>
<thead>
<tr>
<th>Language Background</th>
<th>Control (n = 20)</th>
<th>Trained (n = 95)</th>
<th>Single Talker, Lexical Feedback (n = 20)</th>
<th>Multiple Talkers, Lexical Feedback (n = 24)</th>
<th>Single Talker, Semantic Context Feedback (n = 22)</th>
<th>Multiple Talkers, Semantic Context Feedback (n = 29)</th>
</tr>
</thead>
<tbody>
<tr>
<td>L1 English</td>
<td>Control (n = 16)</td>
<td>Trained (n = 78)</td>
<td>Single Talker, Lexical Feedback (n = 18)</td>
<td>Multiple Talkers, Lexical Feedback (n = 18)</td>
<td>Single Talker, Semantic Context Feedback (n = 19)</td>
<td>Multiple Talkers, Semantic Context Feedback (n = 23)</td>
</tr>
<tr>
<td>L1 Dutch-L2 English</td>
<td>Control (n = 94)</td>
<td>Trained (n = 29)</td>
<td>Single Talker, Lexical Feedback (n = 18)</td>
<td>Multiple Talkers, Lexical Feedback (n = 18)</td>
<td>Single Talker, Semantic Context Feedback (n = 19)</td>
<td>Multiple Talkers, Semantic Context Feedback (n = 23)</td>
</tr>
</tbody>
</table>
2.4. Stimuli

Following Maye et al. (2008), the NSAE accent was created by implementing a set of cross-category pronunciation deviations from Standard American English speech, outlined in Table 2 (sample items provided in Appendix 1). These particular deviations were selected as they are known to present common pronunciation problems for non-native speakers of English (Avery & Ehrlich, 1992) and thus could be considered plausible deviations that listeners might encounter in foreign-accented speech. Moreover, we wanted to include multiple deviation patterns to more closely emulate the experience of listening to naturally-accented speech. When encountering a novel foreign accent, listeners typically must contend with a variety of accented segments that span the gamut of vowels and consonants, containing some patterns that might be familiar to them (from experience with other dialects) and others that are unfamiliar.

2.4.1. Contrast types

Three accent deviation patterns involved phonemes that are contrastive in Dutch, which would typically assimilate to distinct L1 categories (indicated in Table 2 as “Dutch/English contrasts”). These included two vowel contrasts (/l/ → /h/ and /æ/ → /æ/) and one consonant pattern (/z/ → /s/). It is worth noting that certain Dutch varieties do devoice /z/ to /s/ in certain contexts (van de Velde & Gerritsen, 1996), and fricative voicing can be highly variable (e.g., Gussenhoven, 1999).

The remaining three patterns (/æ/ → /e/, /l/ → /h/ and /d/ → /t/ word-finally) are termed “English only” contrasts (Table 2), ones that are fully distinctive in English but whose status in Dutch is less straightforward. Dutch does not differentiate /l/ → /æ/ and /æ/ → /æ/, and as a result, these segments are perceptually confusable for L2 listeners (Cutler, Weber, Smits, & Cooper, 2004). While English possesses these two midfront unrounded vowels, only one vowel exists in that part of the vowel space in Dutch. It is typically transcribed as /æ/; however, its acoustic production is typically lower than English /æ/ and towards English /æ/ (Cutler & Broersma, 2005). Additionally, /l/ is not a sound that exists in Dutch. Perceptually, /l/ is most frequently misidentified as /l/ by Dutch listeners (Cutler et al., 2004); however, /l/ is the most common substitution in production by Dutch speakers (Wester, Gilbers, & Lowie, 2007). Despite these misperceptions and substitutions in production, Dutch listeners do predominantly produce the segment as /l/ (Hanulíková & Weber, 2012). Moreover, certain English varieties (e.g., Irish English) also produce this /l/ → /t/ substitution.

Finally, Dutch de-voices obstruents word-finally, and as such, Dutch speakers would normally not distinguish /d/ → /t/ in word-final position and often de-voice these segments when speaking English (Booij, 1995). In other syllabic positions, however, this contrast otherwise exists in Dutch. Thus, unlike the “Dutch/English” contrasts, the status of these “English only” contrasts is not clearly phonemic in Dutch, and is instead either non-contrastive or neutralized in Dutch. Word frequency information is provided in Appendix 2.

The training stimuli were naturally-produced by 4 phonetically-trained, male native talkers of Standard American English. Stimuli used in the test tasks were produced by a trained talker as well as a fifth male native talker of English (the generalization talker). Following recent work (Weatherholtz, 2015; White & Aslin, 2011), stimuli were natural rather than synthesized (Maye et al., 2008) to better reflect the processing demands and natural variation listeners would experience when dealing with accented speech. Talkers were provided with IPA transcriptions of the items and instructed to produce them as transcribed. To ensure the appropriate pronunciation shifts were produced, a small group of native English listeners (n = 6) performed a two-alternative forced choice task on a subset of the items (127/513, 25%; e.g., hearing [k]rım and indicating whether they heard “cream” or “krım”). Accuracy was very high for all talkers (M = 98.8%), indicating that they produced the artificial accent as intended.

Training materials were a set of 120 intended real words. Here, “intended” real words indicate that these items would be considered real words when pronounced with a Standard American English accent; however, in NSAE, certain items appeared to be non-words (e.g., snake /snek/ → /snek/). Each item only contained one deviation pattern. The 120 training items were divided between the six accent deviation patterns (16 items per consonant deviation pattern, 24 items per vowel deviation pattern). The output of the NSAE accent had one of two possible outcomes: (1) Minimal pair change (real word /bæit/ → /bet/) or (2) Lexicality change (real word /θɜːst/ → /ñɔːst/). In all conditions, 60 items underwent a Minimal pair change, and 60 items underwent a Lexicality change.

Each probe block contained a unique set of 36 items, which included 12 trained pattern NSAE items (e.g., /bæit/ → /bet/), 18 real words, and 6 non-words (e.g., /flɔːt/). A total of 72 different items were included across the two probe blocks. All of the trained pattern items involved a lexicality change (real word /θɜːst/ → /ñɔːst/). None of these items were seen during the training blocks, and they were produced by a trained talker.¹

The lexical decision task presented a total of 189 items: 54 trained pattern NSAE items, half of which were presented during training and half novel, 25 items with untrained patterns, 80 real words, and 30 non-words. Similar to the probe task, trained pattern items involved lexicality changes, such that they would be considered non-words to untrained listeners. Untrained patterns were accent deviation patterns not included during the training blocks but would be considered phonologically/feature-related to the trained patterns (e.g., trained pattern: /z/ → /s/, untrained pattern: /n/ → /t/). Half of the stimuli were produced by a trained talker, and half by a novel, generalization talker. Which items were produced by which talker

---

¹ Due to the constraints of the accent, some items (less than 10% of the 513 total stimulus items) included in the study that were nonwords in English were real words in Dutch. This included two out of 42 nonword fillers and, of the critical items, 12/127 Dutch/English items and 17/119 English only items.
was counterbalanced across participants. In both the probe and lexical decision tasks, real and non-word items did not contain any NSAE-accented segments, only using segments that were unaffected by the NSAE accent. Non-word items, similar to trained and untrained items, were constructed to differ minimally from real words (differing by one phoneme). If all trained tokens were accepted as words, we would expect a 70%/30% split for word/non-word responses. If listeners generalized to untrained accent patterns, then we would expect an 84% lexical endorsement rate. This is consistent with Maye et al. (2008) who utilized 5 types of stimuli and predicted a maximum 80%/20% split following accent exposure.

The word identification task contained a total of 132 items, including 72 trained pattern NSAE items. Half of the trials involved minimal pair changes, and half were lexicality changes. Additionally, half of the items were produced by the talker from the single-talker training condition, and half by the novel, generalization talker. All of these items were not seen in the training blocks or the lexical decision task.

In addition to these tasks, the phonetic assessment task included 30 real words divided evenly between the 6 accent deviation patterns. All words possessed a minimal pair item containing the other segment in the accent pattern (e.g., presenting bed, which is a minimal pair with bad, the other segment in the /ɪd/ ➔/æd/ pattern). This enables us to examine whether these L2 listeners were able to correctly identify items containing these particular sound patterns in Standard American English. All items were produced by a trained talker.

3. Results

3.1. Lexical decision probe task

Lexical endorsement rates (i.e., the number of ‘Word’ responses) were calculated in each of the probe blocks for L1 and L2 listeners. To examine whether talker variability or feedback type during training had an impact on adaptation, binary-coded data (1 = word, 0 = nonword) of responses to critical items were submitted to logistic linear mixed effects regression (LMER) models (Baayen, Davidson, & Bates, 2008) with contrast-coded fixed effects of L1 (Dutch, English), Block (1, 2), Feedback (lexical, semantic context) and Variability (single talker, multiple talker) and their 2- and 3-way interactions. The maximal random effects structure that would converge was implemented, with random intercepts for participant and item, along with random slopes for L1, Feedback and Variability by item and Block by participant. Model comparisons were performed to determine whether the inclusion of each of these fixed effects and their interactions made a significant contribution to the model. Lexical endorsement rates did not significantly differ overall between online and lab participants ($\chi^2 = 1.70, p = 0.192$); thus, the results reported here pool these groups.

As shown in Table 3, significant main effects of Block and L1 were obtained, with a significant increase in lexical endorsement rates overall from Block 1 to Block 2 as well as significantly higher endorsement rates across blocks by L2 listeners relative to L1 listeners. Additionally, significant 2-way interactions were found, including L1 × Block and Feedback × Block. Subsequent models revealed that participants who received lexical disambiguating information during training provided significantly higher lexical endorsement rates for critical items by the second block ($M = 44−71\%$) relative to those who received semantic context information ($M = 47−67\%$). No other effects or interactions reached significance ($\chi^2 < 2.96, p > 0.09$).

To further examine the potential influence of linguistic experience on adaptation during training, a model breaking down the different items by contrast type was also conducted. The proportion of word responses by contrast type and language background is depicted in Fig. 2. This model contained contrast-coded fixed effects of L1 (Dutch, English) and Block (1, 2), and Helmert contrast-coded fixed effects for Item Type (A: Non-word + English only vs. Dutch/English; B: Non-word vs. English only). For additional comparisons within Item Type, models containing Item Type C: Non-word vs. English only + Dutch/English and D: English only vs. Dutch/English were also constructed. Accordingly, the critical $p$ value was set to 0.025. Random intercepts were included for participant and item, along with random slopes for Language (by item and Block and Item Type by participant).

A summary of the significant effects is provided in Table 4. Crucially, as predicted, multiple 3-way L1 × Block × Item Type interactions were significant. Subsequent LMER analyses revealed that while endorsement rates for L1 listeners to Non-word and Trained Patterns (both contrast types) items increased from Block 1 to Block 2, the magnitude of increase was significantly larger for Trained Pattern items ($\chi^2 = 7.69, p = 0.006$). L2 listeners increased their endorsement rates between blocks for Dutch/English items to a greater extent than to Non-words and English only items ($\chi^2 = 5.38, p = 0.02$). No block difference was found in the magnitude of difference between Non-word and English only items ($\chi^2 = 0.002, p = 0.96$) for the L2 listeners.

Comparing L1 and L2 listeners, significantly higher lexical endorsement rates were found for English only and Nonword item types in Block 1 by Dutch listeners relative to English listeners ($\chi^2 > 12.93, p < 0.0003$), though there were no group differences for Dutch/English items. By Block 2, English and Dutch listeners had comparable lexical endorsement rates by item type ($\chi^2 < 1.23, p > 0.27$).

We also examined listeners’ performance in the probe task by individual deviation pattern for the accented items. A 3-way mixed ANOVA was conducted with Language (L1, L2) as a between-subject factor and Block (1, 2) and Accent Pattern (/ɪd/ ➔/u/, /ɪ/ ➔/æ/, /ɪ/ ➔/æ/), /ɪ/ ➔/ui/, /ɪ/ ➔/ui/) as repeated measures. Results revealed significant main effects of Language [F(1, 169) = 29.36, $p < 0.0001$], Block [F(1, 169) = 155.61, $p < 0.0001$], and Accent Pattern [F(5, 169) = 61.43, $p < 0.0001$], as well as a significant 2-way Language × Accent Pattern [F(5, 169) = 23.74, $p < 0.0001$] interaction. A 3-way Language × Block × Accent Pattern interaction was also obtained [F(5, 169) = 2.86, $p = 0.14$]. Follow-up Bonferroni-adjusted pairwise comparisons for each Accent Pattern with Language as a factor in the first block (prior to accent exposure) found significantly higher endorsement rates for L2 listeners than L1 listeners for all contrasts ($p < 0.008$), with the exception of /ɪ/ ➔/ui/ ($p = 0.28$). The same analyses performed on the second probe block (after one block of accent exposure) found higher endorsement rates for L2 than L1 listeners for /ɪd/ ➔/u/
Table 3
Statistical results for the lexical decision probe task: coefficient estimates, standard errors of the coefficients, along with chi-square and p-values for the log-likelihood comparisons for each subset model relative to the full model. For brevity, only the significant effects have been listed.

<table>
<thead>
<tr>
<th>Fixed effects</th>
<th>β</th>
<th>SE β</th>
<th>χ²</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Block</td>
<td>1.20</td>
<td>0.10</td>
<td>108.41</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>L1</td>
<td>-0.70</td>
<td>0.27</td>
<td>6.27</td>
<td>0.012</td>
</tr>
<tr>
<td>L1 × Block</td>
<td>1.21</td>
<td>0.20</td>
<td>33.89</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Feedback × Block</td>
<td>0.42</td>
<td>0.20</td>
<td>4.52</td>
<td>0.03</td>
</tr>
</tbody>
</table>

(p < 0.0001), but significantly higher word responses for L1 than L2 listeners for /l/ → /t/ (p = 0.021) and /i/ → /ɪ/ (p = 0.023). No other contrasts significantly differed by language background (p > 0.59).

To summarize, these results suggest that the type of accent deviation pattern, whether it contained Dutch/English or English only contrasts, interacted with listeners’ linguistic experience to influence the degree to which they accepted accented items as being real words. L2 listeners had an overall higher rate of lexical endorsements than L1 listeners across blocks. The L1 group demonstrated a significant increase in endorsement rates after the first training phase to both non-word and trained pattern items, but critically, a significantly larger endorsement rate increase for trained pattern items. L2 listeners saw a significant increase from Block 1 (prior to training) to Block 2 in lexical endorsement rates for items containing Dutch/English contrasts though not for English only items, presumably due to their relatively high endorsement rates for English only items in Block 1. A comparison of L1 and L2 groups found that prior to training, Dutch listeners endorsed non-words and items containing English only contrasts significantly more than English listeners (with comparable endorsement rates between the two groups for Dutch/English contrast items). By Block 2, both groups had comparable endorsement rates by item type.

3.2. Lexical decision test task

To examine how trained listeners’ endorsement rates compared to control listeners (Fig. 3), a logistic LMER model was constructed with binary-coded word responses (1 = word, 0 = non-word) to trained pattern and non-word items as the dependent variable with fixed effects of L1 (Dutch, English), Training (Control, Trained), and Helmert contrast-coded fixed effects for Item Type (A: Non-word + English only vs. Dutch/English; B: Non-word vs. English only). Additional models were constructed for further within-Item Type comparisons (C: Non-word vs. English only + Dutch/English, D: English only vs. Dutch/English), with the critical p value set to 0.025 to account for multiple comparisons.
for multiple comparisons. Random intercepts for participant and item were included, along with random slopes for L1 and Training by item and Item Type by participant. A summary of the significant statistical effects is provided in Table 5. Again, lexical endorsement rates did not significantly differ overall between online and lab participants ($\chi^2 < 1.75, p > 0.186$) and are thus pooled in the reported analyses. Models examining the factors of talker variability and feedback type during training were also run; however, as those factors and their interactions, did not reach significance ($\chi^2 < 3.34, p > 0.07$), we do not report those models in detail here.

As shown in Table 5, a significant L1 x Training was obtained, and follow-up LMER models revealed significantly higher lexical endorsement rates by L2 control listeners relative to L1 control listeners ($p = 0.004$), but no significant difference between L1 and L2 trained listeners ($p = 0.996$). Training x Item Type interactions reflect significantly higher endorsement rates by trained listeners than control listeners to all item types, including non-words, but a larger group difference for English only and Dutch/English items relative to non-words. This indicates that regardless of linguistic background, listeners were more willing to accept accented items containing trained accent deviation patterns as being words following exposure to NSAE-accented speech.

There were also significant L1 x Item Type interactions. Follow-up analyses revealed that L2 listeners produced significantly higher endorsement rates to English only items relative to Dutch/English items ($p = 0.001$), with no difference in endorsement rates to these items for L1 listeners ($p = 0.22$). Both L1 and L2 listeners endorsed English only items as being words significantly more than non-words ($p < 0.001$); however, the magnitude of this difference was larger for L2 listeners as compared to L1 listeners. The remaining main effects and 3-way interactions did not reach significance ($\chi^2 < 2.96, p > 0.09$).

Analyses were conducted to examine listeners' performance in this test task by individual deviation pattern for the accented items (Fig. 4). A 3-way mixed ANOVA was conducted with Language (L1, L2) and Condition (trained, control) as between-subject factors and Accent Pattern (/d/ → /t/, /i/ → /æ/ → /æl/, /l/ → /l/, /l/ → /l/, /æl/ → /æl/, /l/ → /s/) as repeated measures. Results revealed significant main effects of Language ($F(1, 206) = 19.30, p < 0.0001$), Condition ($F(1, 206) = 65.64, p < 0.0001$), and Accent Pattern ($F(5, 206) = 34.35, p < 0.0001$), as well as significant 2-way Language x Accent Pattern ($F(5, 206) = 4.43, p = 0.001$) and Condition x Accent Pattern ($F(5, 206) = 2.34, p = 0.04$) interactions. No 3-way interaction was obtained ($p = 0.90$). Follow-up Bonferroni-adjusted pairwise comparisons for each Accent Pattern with Language as a factor found significant differences in endorsement rates between L1 and L2 listeners for /d/ → /t/ ($p = 0.007$), /l/ → /æl/ ($p = 0.001$) and /l/ → /l/ ($p = 0.038$). No significant language differences were found for the remaining accent patterns ($p > 0.073$).

Interestingly, the three patterns where significant differences in language background were found were classified as English only in the present work, while the non-significant differences were for the Dutch/English patterns. Overall for L2 listeners, endorsement rates for items with /l/ → /l/ and /æl/ → /l/ were endorsed significantly less than all other contrasts ($p < 0.002$) but did not differ from each other ($p = 0.122$). Endorsement rates for items with /l/ → /t/ and /l/ → /s/ were intermediate relative to the other contrasts.

Moreover, to examine whether multi-talker training influenced performance on the trained versus generalization talker, a model with contrast-coded fixed effects of Speaker (trained, generalization), Talker Variability (single, multiple), and Item Type (critical item, nonword) was constructed for the trained listeners, with random slopes for Talker Variability by item and Speaker and Item Type by subject. In addition to the expected significant effect of Item Type ($\chi^2 = 43.71, p < 0.0001$), the only other effect to reach significance was Speaker ($\beta = 0.18$, SE $\beta = 0.05$, $\chi^2(1) = 11.07, p = 0.0009$), with higher endorsement rates overall for the trained relative to generalization speaker. Critically, Speaker did not interact with either factor, indicating that listeners were able to generalize from the trained talker(s) during exposure and that the number of talkers to which they were exposed did not affect their ability to generalize.

In sum, trained listeners endorsed items containing trained patterns significantly more than non-words relative to control listeners. L2 control listeners had significantly higher endorsement rates overall relative to L1 control listeners; however, no significant difference emerged between L1 and L2 trained groups. Furthermore, the difference in endorsement rates to English only items relative to non-word items was significantly larger for L2 versus L1 groups.

### 3.3. Word identification task

Two types of stimuli were included in the word identification task: minimal pair (where the item would be considered a word in a Standard American English accent and a different word in NSAE) and lexicality change (where the item would be considered a non-word in Standard American English but a word in NSAE). Responses to minimal pair items were binary-coded in two ways: (1) identification accuracy in NSAE, termed NSAE Accuracy, and (2) identification accuracy in a Standard American English accent, termed SAE Accuracy. For instance, the word “pod” would be produced as [pæld] “pot” in NSAE. If listeners transcribed it as “pod”, then this would be correct in terms of SAE Accuracy (i.e., they did not consider that the item’s NSAE pronunciation could reflect the intended word “pod”). Conversely, if they transcribed “pod”, then their NSAE Accuracy would increase but not their SAE Accuracy.

For lexicality change items, NSAE identification accuracy was also calculated. This would involve cases where, for example, listeners were presented with an item produced as [ʃælf] and accurately transcribed it as “shelf”. The number of non-word responses (denoted by an ‘X’ by participants) was determined for both minimal pair and lexicality change items; however, the results here focus on the lexicality change data (as the proportion of minimal pair items identified as non-words was relatively low, $M = 7$%).

To compare word identification performance between L1 and L2 listeners, a logistic LMER model on NSAE Accuracy for lexicality change items with trained patterns was constructed (Fig. 5), with fixed effects of L1 (Dutch, English), Training (Trained, Control) and Contrast Type (English only, Dutch/English), along with their interactions. Random intercepts for
participant and item were included, as well as by-item random slopes for L1 and Training, as well as a by-participant slope for Contrast Type (significant effects provided in Table 6). Overall NSAE and SAE Accuracy rates did not significantly differ between online and lab participants ($\chi^2 < 2.16$, $p > 0.14$). Similarly, models run with talker variability and feedback type as factors did not find these effects or their interactions to be significant for NSAE and SAE Accuracy rates ($\chi^2 < 1.59$, $p > 0.21$); thus, the results reported here pool these groups.

A significant effect of Training was obtained, indicating that across L1 backgrounds, participants who underwent training were more likely to accurately identify words based on the NSAE accent relative to control listeners (that is, in Fig. 5, the darkest portion of the bar is smaller in the control relative to the trained group). 2-Way Training $\times$ L1 and Contrast Type $\times$ L1 interactions were found. Moreover, a marginally significant 3-way Training $\times$ Contrast Type $\times$ L1 interaction was obtained. Subsequent analyses revealed that L2 control listeners had significantly higher NSAE Accuracy rates for English only items than L1 control listeners ($\chi^2 < 1.59$, $p > 0.21$); thus, the results reported here pool these groups.

A significant effect of Training was obtained, indicating that across L1 backgrounds, participants who underwent training were more likely to accurately identify words based on the NSAE accent relative to control listeners (that is, in Fig. 5, the darkest portion of the bar is smaller in the control relative to the trained group). 2-Way Training $\times$ L1 and Contrast Type $\times$ L1 interactions were found. Moreover, a marginally significant 3-way Training $\times$ Contrast Type $\times$ L1 interaction was obtained. Subsequent analyses revealed that L2 control listeners had significantly higher NSAE Accuracy rates for English only items than L1 control listeners ($\chi^2 < 1.59$, $p > 0.21$); thus, the results reported here pool these groups.

A significant effect of Training was obtained, indicating that across L1 backgrounds, participants who underwent training were more likely to accurately identify words based on the NSAE accent relative to control listeners (that is, in Fig. 5, the darkest portion of the bar is smaller in the control relative to the trained group). 2-Way Training $\times$ L1 and Contrast Type $\times$ L1 interactions were found. Moreover, a marginally significant 3-way Training $\times$ Contrast Type $\times$ L1 interaction was obtained. Subsequent analyses revealed that L2 control listeners had significantly higher NSAE Accuracy rates for English only items than L1 control listeners ($\chi^2 < 1.59$, $p > 0.21$); thus, the results reported here pool these groups.

A significant effect of Training was obtained, indicating that across L1 backgrounds, participants who underwent training were more likely to accurately identify words based on the NSAE accent relative to control listeners (that is, in Fig. 5, the darkest portion of the bar is smaller in the control relative to the trained group). 2-Way Training $\times$ L1 and Contrast Type $\times$ L1 interactions were found. Moreover, a marginally significant 3-way Training $\times$ Contrast Type $\times$ L1 interaction was obtained. Subsequent analyses revealed that L2 control listeners had significantly higher NSAE Accuracy rates for English only items than L1 control listeners ($\chi^2 < 1.59$, $p > 0.21$); thus, the results reported here pool these groups.

A significant effect of Training was obtained, indicating that across L1 backgrounds, participants who underwent training were more likely to accurately identify words based on the NSAE accent relative to control listeners (that is, in Fig. 5, the darkest portion of the bar is smaller in the control relative to the trained group). 2-Way Training $\times$ L1 and Contrast Type $\times$ L1 interactions were found. Moreover, a marginally significant 3-way Training $\times$ Contrast Type $\times$ L1 interaction was obtained. Subsequent analyses revealed that L2 control listeners had significantly higher NSAE Accuracy rates for English only items than L1 control listeners ($\chi^2 < 1.59$, $p > 0.21$); thus, the results reported here pool these groups.

A significant effect of Training was obtained, indicating that across L1 backgrounds, participants who underwent training were more likely to accurately identify words based on the NSAE accent relative to control listeners (that is, in Fig. 5, the darkest portion of the bar is smaller in the control relative to the trained group). 2-Way Training $\times$ L1 and Contrast Type $\times$ L1 interactions were found. Moreover, a marginally significant 3-way Training $\times$ Contrast Type $\times$ L1 interaction was obtained. Subsequent analyses revealed that L2 control listeners had significantly higher NSAE Accuracy rates for English only items than L1 control listeners ($\chi^2 < 1.59$, $p > 0.21$); thus, the results reported here pool these groups. A model containing the same fixed and random effects structure was constructed with SAE Accuracy on minimal pair change items as the dependent variable (Table 8; Fig. 6). Results revealed significant main effects of Training and L1, along with a significant Training $\times$ L1 interaction. Subsequent LMER models found that L1 control listeners produced significantly higher NSAE Accuracy rates than L2 trained listeners ($p = 0.01$).
The effect of training was significant for both L1 and L2 groups found a significant main effect of Training \( F(1, 203) = 24.23, p < 0.0001 \) and Accent Pattern \( F(5, 203) = 52.65, p < 0.0001 \) were found, along with a significant Language \( \times \) Accent pattern interaction \( F(1, 203) = 8.30, p < 0.0001 \). All other effects and interactions were not significant (\( p > 0.242 \)).

Bonferroni-adjusted pairwise comparisons for each Accent Pattern with Language as a factor found that L2 participants had significantly higher NSAE Accuracy for items with /d/ \( \rightarrow /L/ (p < 0.0001) \) and /s/ \( \rightarrow /æ/ (p = 0.011) \) patterns relative to L1 listeners. Conversely, L1 participants produced higher NSAE Accuracy rates than L2 listeners for items with /L/ \( \rightarrow /L/ (p = 0.05) \) and /æ/ \( \rightarrow /s/ (p = 0.031) \). Overall, for L2 listeners, items with /L/ \( \rightarrow /L/ \) and /æ/ \( \rightarrow /s/ \) had significantly lower NSAE Accuracy rates relative to all other contrasts (\( p < 0.0001 \)). Performance on the other contrasts did not differ from each other (\( p > 0.103 \)).

Finally, to examine whether multi-talker versus single-talker exposure would affect listeners’ word identification performance and generalization to a novel talker, a model on NSAE Accuracy rates was constructed with contrast-coded fixed effects of Talker Variability (single, multiple) and Speaker (trained, generalization), with random slopes of Talker Variability by item and Speaker by participant. The only significant effect was Speaker \( (β = −0.26, SE \ β = 0.10, χ^2(1) = 6.22, p = 0.013) \), with higher NSAE accuracy rates for the generalization talker as compared to the trained talker. Speaker, however, did not interact with Talker Variability, indicating that performance on trained and generalization talkers was not affected by the number of talkers received during training.

The results of the word identification task provide converging evidence with the lexical decision test task of the influence of accent exposure and linguistic background on adaptation. For items that would be considered non-words in Standard American English (lexicibility change items), trained listeners, across language backgrounds, were more likely to accurately identify words based on the NSAE accent as compared to control listeners. That is, upon hearing the item [kæk], trained listeners were more likely to identify it as “cake”. Both L1 and L2 trained listeners identified these items as non-words significantly less than control listeners (though the effect of training was significantly larger for L1 listeners). For both lexicibility change and minimal pair items, L2 control listeners had higher NSAE accuracy rates for English only items relative to L1 control listeners, with comparable performance by trained listeners. For Dutch/English contrasts, control listeners from both language groups performed comparably, and L1 trained listeners had higher accuracy NSAE rates than L2 trained listeners. Moreover, SAE accuracy rates for minimal pair items were higher for L1 control versus L1 trained listeners. This indicates that upon hearing the item [apot], control listeners were more likely to identify the item as “pot”. This group difference did not reach significance for L2 listeners.

### 3.4. Phonetic assessment task

To determine whether L2 listeners differed with respect to their ability to identify SAE-accented contrasts, word identification accuracy was tabulated for the phonetic assessment task (Fig. 8), which they completed after training and test tasks. A

---

**Fig. 4.** Mean proportion of word responses in lexical decision test task for trained pattern items by individual accent pattern for L1 (top panel) and L2 (bottom panel) listeners by Training (control, trained). Errors bars denote ±1 standard error.

Numerical trend in the appropriate direction. These findings indicate that listeners who did not receive exposure to the NSAE accent were more likely to identify minimal pair change items (e.g., the word “pod” produced as [pot] “pot”) as its surface form (e.g., “pot”).

Finally, an analysis of non-word response rates for L1 and L2 groups found a significant main effect of Training along with significant 2-way interactions of Contrast Type \( \times \) and Training \( \times \) L1 (Table 9). Follow-up analyses indicate that while the magnitude of this difference was significantly larger for L1 listeners as compared to L2 listeners.

Similar to the lexical decision test analyses, we also examined NSAE Accuracy for lexicibility change items broken down by individual accent pattern (Fig. 7). A 3-way mixed ANOVA was conducted with Language (L1, L2) and Training (control, trained) as between-subjects factors and Accent Pattern (/d/ /t/, /s/ → /æ/, /f/ → /L/, /fl/ → /L/, /æsl/ → /æ/, /z/ → /s/) as repeated measures. Significant main effects of Training \( F(1, 203) = 24.23, p < 0.0001 \) and Accent Pattern \( F(5, 203) = 52.65, p < 0.0001 \) were found, along with a significant Language \( \times \) Accent pattern interaction \( F(1, 203) = 8.30, p < 0.0001 \). All other effects and interactions were not significant (\( p > 0.242 \)).
logistic LMER was constructed with accuracy as the dependent variable (1 = correct, 0 = incorrect) and contrast-coded fixed effects of Training (Control, Trained) and Contrast Type (Dutch/English, English only), with participant and item as random intercepts, and a by-participant random slope for Contrast Type and a by-item random slope for Training. No significant effects or interactions were found ($\chi^2 < 2.48$, $p > 0.12$). Though, there was a numerical trend for words containing Dutch/English contrasts to be more accurately identified ($M = 79\%$) relative to words with English only contrasts ($M = 69\%$). L2 listeners’ accuracy rates were compared against a group of L1 listeners who completed this task on Amazon Mechanical Turk ($n = 20$), and a significant effect of L1 was found ($p < 0.001$), whereby native listeners were significantly more accurate at identifying these words ($M = 92\%$) relative to the L2 listeners ($M = 75\%$).

4. Discussion

One of the aims of the present work was to examine whether perceptual adaptation entails accent pattern-specific adjustments or a more general relaxing of input-to-representation mappings across the system. If as a result of accent exposure, listeners were generally relaxing their criteria for permissible input-to-representation mappings, this would be evident in their lexical endorsement rates, with comparable endorsement rates for both trained pattern items and non-words, as they would be more willing to accept all types of items as being possible words. In contrast, specificity of learning would manifest as higher lexical endorsement rates for trained pattern items relative to non-words by the end of training, as this would reflect training-induced, pattern-specific adjustments. The current findings revealed that exposure to NSAE, containing an array of both vowel and consonantal deviations from Standard American English, yielded an overall increase in native listeners’ willingness to accept non-word forms as being possible English words, as indicated by the significantly higher lexical endorsement rate of non-word items for trained as compared to control listeners. This suggests that exposure to an unfamiliar accent involves a general relaxing of criteria for what counts as an acceptable match between stored lexical representations and the incoming speech input.
This is consistent with what has been found in cases where listeners are in more adverse listening conditions, such as listening in noise or to reduced speech (Brouwer, Mitterer, & Huettig, 2012; McQueen & Huettig, 2012). In this case, listeners who were exposed to productions that deviated strongly from Standard American English norms, and who received either direct orthographic or semantic–contextual information about the nature of the deviations, increased their overall tolerance for mismatches between input and representation.

However, this mismatch tolerance appeared to be constrained, as lexical endorsement rates and word identification NSAE accuracy were still significantly higher for items containing trained accent patterns relative to non-words. Thus, participants had learned to distinguish between items that deviated from expected word production due to the expression of the trained accent and items that did not match any expected word productions (i.e., items that were simply not English words). It is important to note that the non-words employed in the present work were minimally different from real words (in that they differed by a single phoneme). Prior work with this experimental paradigm has utilized “maximal non-words” (Weatherholtz, 2015), which differed from real words on multiple segments and features, and found listeners were willing to endorse items consistent with the back vowel lowered accent to which they were exposed and reject maximal non-words as being possible words. The “minimal non-words” used in this study provided a more sensitive test of the system’s ability to differentiate exposed non-accent patterns from a general relaxing of criteria for non-word items. If non-word items were distinctly non-word-like (i.e., differed from real words by more than just one segment), listeners would presumably be less likely to tolerate or consider them to be words, even if their criteria for word status had been somewhat relaxed. Indeed, these minimal non-words resulted in higher lexical endorsement rates for both L1 and L2 listeners as compared to Weatherholtz (2015) but are consistent with prior work similarly using “near words” or minimal non-words (e.g., Cutler & Broersma, 2005).

The current findings indicate that while the perceptual system was generally increasing its baseline tolerance for atypical speech input, it was still constraining this tolerance to focus on adjusting distributions for specific categories. L1 listeners in the present work did not loosen their criteria to accept, to the same degree as the trained pattern items, minimal non-words such as “spaish” [spaj] or “spum” [spum], which, if the L1 listeners had relaxed their lexical criteria more pervasively, could have potentially been recognized as “spice” or “spun”, respectively. This influenced not only the perception of non-words but also existing real words. For instance, when presented with an accented production [put], L1 trained listeners were more likely to entertain it as a possible production of “pod” than L1 control listeners. L1 control listeners who had...
not received NSAE-accented exposure would possess prior beliefs about category distributions based on their experience with Standard AE-accented speakers that would have led them to, for example, strongly activate the word *pot* and send relatively less activation to *pod*. Trained L1 listeners, as a result of bottom-up exposure and top-down information (lexical or semantic contextual feedback), updated their beliefs about the distributions of the relevant categories when encountering this accent, which in this example actually collapses two categories, sending activation to both *pot* as well as *pod*.

The current study also investigated how L2 listeners accommodate pronunciation variation involving contrasts that both
exist and do not exist in their second language. The findings of the present work demonstrated enhanced flexibility within the perceptual system for L2 listeners, as evidenced by overall higher endorsement rates from L2 control listeners and trainees prior to training relative to L1 listeners. That is, Dutch-English listeners, without any training or exposure to the NSAE accent, were more likely to consider non-word items as being words in English than native English listeners. It is conceivable that when listening in a second language, where listeners are aware that their linguistic knowledge is less robust or as a result of experience with juggling more than one language (Weber et al., 2014), they may be more willing to be flexible in their phonetic-to-lexical mappings. Crucially, while L2 listeners’ initial baseline tolerance for mismatched mappings might have been higher relative to L1 listeners, both L1 and L2 listeners exhibited training-induced, pattern-specific adjustments.

The specificity of perceptual adaptation was also highlighted by the fact that adaptation was modulated by the type of phonemic contrast employed in the accent deviation pattern, that is, its phonemic status in the L2 listeners’ native language. Three of the deviation patterns involved segments whose phonemic status in Dutch is less straightforward, with segments that either assimilate to a single category (/sl/ – /æl/, /nl/ – /nt/) or are neutralized in a particular context (/dl/ – /nt/ word-finally) in Dutch, which were termed “English only” contrasts. The other three patterns were considered to be distinctive in Dutch (/hl/ – /h/; /l/ – /l/; /l/ – /sl/), referred to as “Dutch/English” contrasts. It should be noted that there were word frequency asymmetries between English only and Dutch/English items (Appendix 2). However, for certain tasks, Dutch/English items were more frequent (lexical decision task and word identification) while in others (probe lexical decision) less frequent. Despite this, the pattern of results across tasks remained consistent across tasks, suggesting that frequency was not the driving force behind this pattern. Overall, the present study found that items containing Dutch/English contrasts were less likely to be considered words prior to training (and by L2 control listeners) than those with English only contrasts.

This difference could be attributed to L2 listeners being more familiar with the variant pronunciations associated with items containing English only contrasts. Dutch-English listeners are more likely to encounter a Dutch-accented English talker producing [tʃstl] for thirst than [krmpl] for cream. As such, this familiarity with the specific cue distributions associated with Dutch-accented English may have informed their beliefs about the specific generative linguistic model for this NSAE accent, increasing their willingness to accept [tʃstl] and other items with English only contrasts as being English words, even without any accent exposure. Though it is worth noting that the set of English only contrasts used in this study is not completely homogeneous with respect to their status in Dutch. For example, the /sl/ – /æl/ distinction is collapsed to a single category in Dutch, while the /dl/ – /nt/ contrast is only neutralized word-finally. As a result, Dutch listeners have no difficulty perceptually distinguishing /d/ and /nt/, while /sl/ – /æl/ is considerably more challenging. This may mean that they maintain distinct /dl/ – /nt/ categories, while their /sl/ – /æl/ categories are less distinct. L2 listeners behaved similarly on these different contrasts (that is, producing higher endorsement rates with little evidence of adaptation), but the reasons underlying why these results arose may differ as a function of the specific contrast. If we consider the /nl/ → /nt/ and /dl/ → /nt/ (word-finally) accent patterns, these were both patterns that L2 listeners would have been familiar with from Dutch-accented English. In the context of the ideal adapter framework (Kleinschmidt & Jaeger, 2015), L2 listeners likely held preliminary beliefs about the distributions of these categories, based on their own prior experience with pronunciation variants of English along with their native language influences on the shape of these distributions that would have led them to initially categorize English only items as words. Thus, no additional adjustments to the distributions would need to be made. On the other hand, the high endorsement rates for the pattern /sl/ → /æl/ may have arisen as a result of L2 listeners possessing less distinct categories for those specific phonemes (e.g., /cutl/ & /broersma/, 2005). Indeed, performance was poorest on the phonetic assessment task for that particular contrast. This perceptual imprecision may have led to the activation of minimally-related items (e.g., /wɛst/ upon hearing [wæst]), leading to enhanced endorsement rates. It remains for future work to disentangle these different potential underlying influences on perceptual adaptation with L2 listeners.

While their initial lexical endorsement rates were high, adaptation was not evident for items with English only contrasts for the L2 listeners. L2 trained listeners did not improve significantly on these items as a result of training, with no significant increase in endorsement rates during the probe task, and no L2 control versus L2 trained group differences on the test tasks. With starting performance already high for these items, it could be the case that L2 listeners were at a limit for how broad or shifted these specific distributions could become. Indeed, we saw that L1 listeners, in response to NSAE-accented exposure, came to approximate L2 listeners’ performance for English only items, shifting or expanding relevant sound category boundaries to accommodate these atypical exemplars.

L2 trained listeners did show an effect of training for items containing Dutch/English contrasts, with an increase in lexical

![Fig. 8. Proportion word identification accuracy in assessment task by contrast. Light grey bars denote contrasts designated ‘English only’, dark grey bars ‘Dutch/English’.](image-url)
endorsement rates in the probe tasks and higher NSAE Accuracy in the word identification task as compared to the control listeners. Adaptation being evident for these contrasts is consistent with the L1 listeners, who demonstrated adaptation for both contrast types (English only and Dutch/English), which is not surprising seeing as both types are contrasts that exist in English. L2 listeners may have initially held a higher degree of certainty about the shape of the distributions of Dutch/English contrasts. This, in conjunction with relatively less familiarity with variant pronunciations for items involving these contrasts, would have led to their initially lower lexical endorsement rates for these items. NSAE-accent exposure, however, provided sufficient evidence for listeners to adjust their distributions, improving their willingness to accept these variant pronunciations as possible words and their accuracy in identifying the intended words. It is worth pointing out that one particular pattern classified as Dutch/English in the present work (/z/ → /s/) did pattern more similarly to the English only items in certain tasks (see Fig. 7). This may reflect this pattern’s more variable status in both English and Dutch, as a fricative devoicing may be present in some listeners’ dialects. The present findings highlight the importance of including a range of accent patterns to examine adaptation, as performance can differ substantially between patterns.

However, L2 trained listeners were overall less accurate at identifying Dutch/English contrast items based on their knowledge of the NSAE accent relative to L1 trained listeners. It is important to note that their performance in the phonetic assessment task was not completely native-like, as L1 listeners were still found to outperform them at identifying English words, suggesting that these L2 listeners may still possess a degree of linguistic uncertainty when listening in their second language. While prior work (Reinish et al., 2012) has shown a comparable degree of category boundary shifting between L1 and L2 listeners in phoneme categorization tasks following exposure to an ambiguous sound, the present findings reveal that L2 listeners may have a more difficult time utilizing this knowledge at a higher-level of processing (namely, lexical identification). If we consider these findings in the context of uncertainty, this performance discrepancy as a product of language background may have resulted from L2 listeners maintaining an overall higher level of uncertainty when listening to English relative to L1 listeners. As a result, L2 trained listeners were more likely to identify a Dutch/English contrast item as a non-word (in the case of lexicality change items) or identify it based on how it was actually produced (for minimal pair items) relative to L1 listeners. It may be the case that listeners with a higher level of uncertainty about the nature of the relevant generative model, in this case our Dutch-English listeners, would require more evidence (either through additional training or more explicit training) in order to effectively update the beliefs about that model relative to those with lower levels of uncertainty. Moreover, the stimuli in the present study were produced by native speakers of English. One could imagine that if L1 Dutch speakers produced these stimuli (with the same artificial accent markers), the non-artificially accented segments would also be affected. This extra variability and the additional pronunciation changes would likely increase processing load and slow the adaptation process for listeners overall. However, it is possible that in this case, listeners with the same language background as the speaker would have some of this processing load alleviated, as they would be more familiar with the accent variation patterns as a function of sharing the same language background. It remains for future work to investigate the potential influence of non-target pronunciation variability on adaptation to target pronunciation changes.

Interestingly, NSAE-accented exposure actually introduced a degree of ambiguity into word recognition for both L1 and L2 groups, as trained listeners began to not only activate the lexical items (as pronounced) but also their possible NSAE variants, as evidenced by their increased likelihood of identifying minimal pair change items as different words than pronounced (e.g., the item [put] identified as pod following training on the NSAE accent which contains final devoicing). Increased activation of multiple lexical possibilities does introduce ambiguity that is potentially detrimental to individual word recognition (i.e., homophony). However, in more natural communicative contexts, given that listeners are proficient at drawing upon higher-level contextual information to facilitate lexical access, the benefit gained from having constructed an accent-specific model of cue distributions as a result of accent exposure should outweigh this seeming disadvantage. Future work could test this issue by providing listeners with sentence-length materials.

Prior work on accent adaptation has typically only used either vowels (e.g., Maye et al., 2008; Weatherholtz, 2015) or consonants in their artificial accent (e.g., Eisner & McQueen, 2005; Kraljic & Samuel, 2007), with vowel deviation patterns typically yielding robust talker-general adaptation and consonant deviation patterns more limited adaptation success. Weatherholtz (2015) posited that this discrepancy may arise because consonants contribute relatively more to lexical identity and word recognition as compared to vowels. This is likely due to a number of factors including the higher number of consonants relative to vowels in phonemic systems and the fact that consonants appear to more tightly constrain lexical selection than vowels (Cutter, Sebastiani-Gallés, Soler-Vilageliú, & Van Ooijen, 2000; Nespor, Peña, & Mehler, 2003), which may stem from listeners accruing experience with the fact that vowels perceptually vary more in context than consonants. The present study is the first to combine both consonantal and vowel deviation patterns in the same artificial accent. An examination of the individual accent patterns (e.g., Fig. 4) revealed there to be no split in performance based on whether an accent pattern involved vowels or consonants. Rather, the accent pattern’s familiarity to the listener and the status of the relevant phonemes in the listeners’ native language appeared to be a stronger influence on the present findings.

Moreover, there has been the suggestion in previous work that exposure to multiple accented talkers promotes generalization to a novel talker that shares the same accent (e.g., Bradlow & Bent, 2008). The present work, however, found no difference in listeners’ ability to generalize to a novel talker at test as a function of talker variability during training. This is in line with recent work (Weatherholtz, 2015), who similarly reported generalization to a new talker following single-talker accent training. Given that the training and generalization talk-
ers were all males producing categorical shifts involving native language categories, they were likely acoustically similar enough to each other to enable cross-talker generalization (Kraljic & Samuel, 2005; Reinisch & Holt, 2014).

In sum, accent exposure was found to induce both a constrained increase in tolerance for atypical speech input and targeted adjustments to specific categories. The findings of the present work highlight the influence of prior experience, namely linguistic background, on listeners’ ability to flexibly accommodate atypical pronunciations during speech perception. L2 listeners were found to effectively utilize either lexical or semantic contextual information to adapt to a Non-Standard American English accent; however, this adaptation process was mediated by the particular phonemic contrasts employed in the items. Listeners’ perceptual confusions with a challenging English contrast and their prior experience with English pronunciation variants, similar to the ones employed in the NSAE accent, enabled listeners to endorse more items as words and identify them more accurately, even without training (English only contrasts) as compared to items with contrasts that exist in their native language (Dutch/English contrasts). However, training-induced adaptation was only found for Dutch/English contrast items, perhaps because L2 listeners’ cue distributions for English only items were already shifted or broadened as much as they could be and reflected the patterns of the NSAE accent. In order to determine whether L2 listeners could show evidence of adaptation to these particular contrasts, future work could compare exposure to accent patterns involving contrasts that are neutralized in their L1 but involve deviations that are either familiar or unfamiliar to the listener (e.g., /d/ → /l/ and /æ/ → /a/). This could tease apart whether listeners had reached a kind of ceiling when it came to movement of the category distributions as a product of familiarity with Dutch-accented English, which employs similar deviation patterns, or if in fact listeners held greater uncertainty about these particular contrasts (being neutralized in their L1), promoting stability over adaptation.

With English being spoken as a second language by over 600 million people worldwide (Simons & Fennig, 2018), the growing need to communicate across a language barrier (i.e., between native and non-native speakers) is uncontroversial. One could address the potential problem of communication across a language barrier in one of two ways: (1) training non-native speakers to modify their foreign accents, or (2) training listeners to be more flexible in their perceptual accommodation to foreign accents. Training non-native speakers to achieve native-like production targets is a well-documented challenge, sometimes requiring weeks of training to yield improvements on only a single segmental contrast (Hirata, 2004; Thomson & Denning, 2015). However, given that listeners must constantly contend with extensive talker-related variability in their speech input and are in fact remarkably efficient at learning to accommodate this variation, training listeners to better understand foreign-accented speech would appear to be more effective than the alternative. Indeed, the current study has potential implications for the development of more efficient speech perception enhancement procedures for cross-linguistic communication, highlighting the need to take into consideration the language background of the listeners as well as accent familiarity and the phonemic status of the segments in the accent patterns.

Acknowledgments

We would like to thank Chun Liang Chan and Alexandra Saldan for their technical and research support. Thanks also to Matt Goldrick and Nina Kraus for their comments and suggestions. We are grateful to Mirjam Ernestus and her research group at the Max Planck Institute for Psycholinguistics in Nijmegen. This work was supported by the NIH National Institute on Deafness and Other Communication Disorders [R01-DC005794] and a Northwestern University Graduate Research Grant. Portions of this work were presented at the 57th Annual Meeting of the Psychonomic Society in Boston, Massachusetts.

Appendix A

See Appendix 1 and 2

Appendix 1

Sample stimuli for each stimulus type (Lexicality change, Minimal Pair change) and contrast type (Dutch/English, English only), as well as sample non-word items. Left column provides NSAE production (IPA transcribed for non-word items); right column provides intended word (Standard American English).

<table>
<thead>
<tr>
<th>Sample stimuli for each stimulus type (Lexicality change, Minimal Pair change) and contrast type (Dutch/English, English only), as well as sample non-word items. Left column provides NSAE production (IPA transcribed for non-word items); right column provides intended word (Standard American English).</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Lexicality change</strong></td>
</tr>
<tr>
<td>-----------------------</td>
</tr>
<tr>
<td>nöss</td>
</tr>
<tr>
<td>säs</td>
</tr>
<tr>
<td>stät</td>
</tr>
<tr>
<td>bëf</td>
</tr>
<tr>
<td>bïck</td>
</tr>
<tr>
<td>sëm</td>
</tr>
<tr>
<td>course</td>
</tr>
<tr>
<td>zïnc</td>
</tr>
<tr>
<td>chip</td>
</tr>
<tr>
<td>hill</td>
</tr>
<tr>
<td>pepper</td>
</tr>
<tr>
<td>sell</td>
</tr>
<tr>
<td>sëln</td>
</tr>
<tr>
<td>nûk</td>
</tr>
<tr>
<td>foup</td>
</tr>
<tr>
<td>born</td>
</tr>
<tr>
<td>sták</td>
</tr>
</tbody>
</table>

Appendix 2

Lexical frequency of the critical items (without any pronunciation changes) for each task by contrast type (Dutch/English, English only). LC: Lexicality change items; MC: Minimal pair change items. LC items become non-words with the application of the accent and thus do not have lexical frequency. MC items become a different word, whose frequency is listed as “accented item.”

<table>
<thead>
<tr>
<th>Sample stimuli for each stimulus type (Lexicality change, Minimal Pair change) and contrast type (Dutch/English, English only), as well as sample non-word items. Left column provides NSAE production (IPA transcribed for non-word items); right column provides intended word (Standard American English).</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Lexicality change</strong></td>
</tr>
<tr>
<td>-----------------------</td>
</tr>
<tr>
<td>LC: 77.04</td>
</tr>
<tr>
<td>LC: 60.66</td>
</tr>
<tr>
<td>LC: 267.16</td>
</tr>
<tr>
<td>LC: 308.52</td>
</tr>
<tr>
<td>LC: 1616.72</td>
</tr>
<tr>
<td>LC: 123.51</td>
</tr>
<tr>
<td>LC: 321.15</td>
</tr>
<tr>
<td>LC: 123.51</td>
</tr>
<tr>
<td>LC: 306.60</td>
</tr>
<tr>
<td>LC: 1002.36</td>
</tr>
</tbody>
</table>

Appendix 3

See Appendix 1 and 2