

---

JOURNAL ARTICLE

# Do social preferences matter in lexical retuning?

Molly Babel<sup>1</sup>, Brianne Senior<sup>2</sup> and Sophie Bishop<sup>2</sup>

<sup>1</sup> Department of Linguistics, University of British Columbia, CA

<sup>2</sup> School of Audiology and Speech Sciences, University of British Columbia, CA

Corresponding author: Molly Babel ([molly.babel@ubc.ca](mailto:molly.babel@ubc.ca))

---

In perceiving spoken language, listeners not only recognize and comprehend the intended meaning of the speaker's words and phrases; they also assess the social dimensions of the speaker. In the present study, we ask whether perceptual learning – a process by which listeners adapt to novel pronunciations – is modulated by listeners' social preferences. To this end, we use a novel accent exhibiting a back vowel lowering chain shift in pleasant and unpleasant conditions in a lexically guided perceptual learning paradigm to test whether listeners adapt less to the unpleasant guise. Experiment 1 confirms that listeners disprefer the unpleasant guise. Using a lexical decision task as a measure of lexical adaptation, Experiment 2 indicates that listeners in the pleasant and unpleasant guises learned the back vowel shift compared to listeners in a control group. These results suggest that exposure to a voice with lower social prestige does not inhibit lexical adaptation.

---

**Keywords:** voice evaluation; perceptual learning; lexical decision; priming; adaptation

---

## 1. Introduction

Spoken language is a barrage of acoustic-phonetic material. As infants acquire their native language(s), they learn to attend to certain aspects of this multidimensional phonetic signal (Werker & Tees, 1984), and as adults, listeners from different language backgrounds prioritize different parts of redundantly cued linguistic contrasts (Francis & Nusbaum, 2002). The ways in which we attend to spoken language are malleable and at least partially dependent on whether our attention is focused on signal-related properties or overall comprehension (Cutler, Mehler, Norris, & Segui, 1987; McAuliffe & Babel, 2016). This perceptual flexibility is necessarily balanced with stability (e.g., Kleinschmidt & Jaeger, 2015). One of the potential ways in which this flexibility/stability balance is accomplished is through the mechanisms and limits of perceptual learning. Generally, perceptual learning is considered a perceptual change as a result of sensory exposure to a deviant or unexpected signal (for a review see Goldstone, 1998); in the context of speech, perceptual learning is a kind of associative learning where what was previously not recognized as a member of a particular sound category or as an interpretable pronunciation of a particular word is now categorized or recognized as such. Here, we are interested in whether social preferences influence this adaptive process.

Evidence for rapid perceptual learning in speech is found in listeners' adaptation to challenging nonnative accents, which is typically quantified in terms of increased intelligibility (e.g., Bradlow & Bent, 2008; Clarke & Garrett, 2004). As nonnative accents are characterized by multiple acoustic-phonetic deviations from the familiar or local accent, perceptual adaptation to such stimuli involves an adjustment to multiple category mappings via meaningful exposure to naturally produced utterances. The more predictable

the sentences, the more rapidly listeners are able to adapt to nonnative accents and the more intelligible these accents become to listeners (Bradlow & Alexander, 2007; Holt & Bent, 2017). Context facilitates this (re)interpretation of this challenging-to-map acoustic-phonetic variation. The paradigm introduced by Norris, McQueen, and Cutler (2003) to study lexically-guided perceptual learning targets particular phonemes more directly, but similarly hinges on context to guide the mapping of an unfamiliar pronunciation. Norris and colleagues synthesized a fricative ambiguous between /f/ and /s/ and spliced this onto Dutch words that ended in either an /f/ or /s/, exposing listeners to these items in the context of a lexical decision task. In the post-test, listeners who heard the ambiguous fricative in the context of the /f/-biased words expanded their /f/ category to accommodate the ambiguous fricative, while listeners who heard the ambiguous fricative in the /s/-biased case likewise increased the size of their /s/ category. Listeners who heard the ambiguous fricative in the context of nonwords showed no shift in their /s/ or /f/ categories, as there was no linguistic context for how to interpret or associate the ambiguous category. This paradigm has been exploited and the concept replicated and extended extensively (e.g., Eisner & McQueen, 2005; Jesse & McQueen, 2011; Kraljic & Samuel, 2005; Kraljic & Samuel, 2006; Kraljic & Samuel, 2007; McAuliffe & Babel, 2016; Reinisch, Weber, & Mitterer, 2013; Reinisch & Holt, 2014; Scharenborg & Janse, 2013; Zhang & Samuel, 2014).

In the present study, we ask whether perceptual learning is modulated by social preferences. It has been robustly established that listeners use learned associations between linguistic categories and social meaning to categorize the phonetic signal. For example, when presented with a *hood-hud* [hʊd] – [hʌd] continuum, listeners are more likely to categorize more steps along the continuum as *hood* when they believe the speaker to be male because they rescale their threshold between these vowels to account for the lower resonant frequencies produced by a typical male vocal tract (Johnson, Strand, & D’Imperio, 1999). The acoustic-phonetic cues to talker sex and gender often convey sound/size relationships and are thus relatively low level and have parallels across the animal kingdom (Reby et al., 2005; Rendall, Owren, Weerts, & Hienz, 2004; see Munson & Babel, to appear for a review of these issues). But, there is also evidence for more clearly social and thus learned associations. For example, with respect to a sound change in progress in New Zealand English, Hay, Warren, and Drager (2006) demonstrate that listeners are more likely to perceive the speech of apparently younger speakers and speakers of apparently lower socio-economic status as being more advanced in the in-progress NEAR/SQUARE merger; this parallels the community-level patterns which indicate that these vowels in younger and lower socio-economic status (SES) speakers are more likely to be merged than older and higher SES speakers. These results show that listeners’ decisions are influenced by social biases, but is the auditory-to-phonetic mapping affected by this real world social knowledge or do those influences exert themselves at a later point? Zheng and Samuel (2017) provide evidence that social biases are likely post-perceptual decision biases and not veridical perceptual warpings. They do this through a selective adaptation paradigm that uses accentedness rating tasks with a continuum of speech samples from native to nonnative. Zheng and Samuel argue that results where listeners rate English words or sentences produced by a speaker with an Asian face as more accented than items produced by a speaker with a White face are the result of what they call *interpretation* as opposed to *perception*.

One way to determine how social meaning affects perception and recognition is to test how social preferences affect what listeners retain from the speech signal. Listeners do not weight all incoming phonetic information equivalently (Johnson, 1997; Sumner, Kim, King, & McGowan, 2014); familiar accents, for example, benefit from improved

encoding—defined by Clopper and colleagues as the “updating [of] the cognitive lexical representation to reflect the current token”—compared to less familiar accents (Clopper, Tamati, & Pierrehumbert, 2016, p. 87). These effects of improved processing for familiar voices and accents are well documented (i.e., the language-familiarity effect; e.g., Goggin, Thompson, Strube, & Simental, 1991; Perrachione & Wong, 2007; Thompson, 1987; Winters, Levi, & Pisoni, 2008). Thus, hearing a familiar voice or accent gives that auditory item a boost compared to a less familiar one. Independent of experience, are speech signals handled differently by listeners? While there have been proposals that listeners attend less to socially dispreferred accents (Lippi-Green, 1997), there is little evidence that the mechanism behind these effects is attention rather than decision biases. There is evidence from the visual system, however, that perceivers are able to essentially spotlight positively- or negatively-valenced emotionally salient images, enhancing the initial processing of the image (Todd, Talmi, Schmitz, Susskind, & Anderson, 2012), which would support a potential role for attention in auditory processing. To test the hypothesis that social preferences affect perception, we used a novel accent with no preconceived social meaning other than a generic social dispreference. A novel accent was necessary in order to disentangle a listener’s social preferences about an accent from their familiarity with that accent, as familiar-sounding speakers are socially preferred (Babel & McGuire, 2015). Sumner et al. (2014) theorize about the relationship between social preferences and familiarity. They argue that, for example, infrequent forms (e.g., words with final voiceless coronal stops that are fully released) are often remembered and processed better than much more frequent forms (e.g., the same words with unreleased or glottalized final /t/) and that such patterns arise as the result of increased attention to socially salient and idealized productions.

To this end we replicate and extend the methods of Weatherholtz (2015), who refined the original methods deployed by Maye, Aslin, and Tanenhaus (2008). Maye and colleagues used a speech synthesizer to manipulate English front vowels into a lowering chain shift (e.g., *witch* as [wɛtʃ] not the local [wɪtʃ]), exposing listeners to this shift in the context of a 20-minute passage from *The Wizard of Oz*. Listeners completed a pre-test and post-test lexical decision task that included critical front vowel items on the same voice used in the read passage. Prior to exposure, listeners did not endorse shifted front vowel words as real words, but after hearing the vowel shift in the context of the story, listeners endorsed the items as words. Novel words which participated in the shift, but were not in the story itself also were endorsed as words, illustrating the chain shift was generalized. A control experiment illustrated that listeners did not globally loosen their word criteria, as testing listeners on a vowel raising chain shift (*witch* as [wɪtʃ]) after being exposed to the vowel lowering shift did not show an increase in word endorsement rates. Yet, the within-subjects design of Maye et al.’s task concerned Weatherholtz (2015), as repeated exposure to nonwords could have led to an increase in word endorsement rates (Zeelenberg, Wagenmakers, & Shiffrin, 2004). Weatherholtz was also concerned with the synthetic nature of the voice used in Maye et al.’s experiment, given the evidence that listeners do not process synthetic and natural speech equivalently (Francis, MacPherson, Chandrasekaran, & Alvar, 2016; Lattner et al., 2003; Smither, 1993; White, Rajkumar, Ito, & Speer, 2009). Thus, Weatherholtz designed a similar task using an adaptation of *The Adventures of Pinocchio* using a between-subjects design with a single lexical decision post-test and a naturally produced chain shift. Weatherholtz also tested how much exposure was necessary to promote learning and generalization across talkers, whereas Maye et al.’s study used the same synthetic voice in training and test. His dissertation reports a series of six experiments probing these issues, with currently germane details being that listeners robustly learned and generalized to passages that varied between approximately 20, six,

and two minutes in length. While learning and generalization was nearly equivalent for the 20- and six-minute conditions, listeners in Weatherholtz's study showed weaker patterns of generalization in the two-minute condition. For this experiment we capitalize on the robust learning that was found in the six-minute story, and use this version in our experiments described below. Note that with our focus on social preferences, our study does not examine listeners' generalization of a learned pattern to a novel speaker, but rather trains and tests lexical retuning within a single speaker, like the original study by Maye and colleagues. If social preferences do influence perceptual learning, we predict that listeners should demonstrate less perceptual learning when exposed to a social dispreferred voice than when their model's voice is more socially preferred.

## 2. Accent and guise creation

A female actor-phonetician produced a story in three guises: (i) a Pleasant Unshifted control condition with standard pronunciations in a pleasant reading voice; (ii) a Pleasant Shifted condition where back vowels are shifted as though participating in a vowel lowering or (F1 raising) chain shift; and, (iii) an Unpleasant Shifted condition where the back vowels are shifted, but the story is produced with a relatively monotone intonation pattern and creaky voice quality. Like the work by Maye and colleagues (Maye et al., 2008) and Weatherholtz (2015), we chose to use a multi-sound chain shift as opposed to a manipulation that involved a single sound or set of sound categories to better mimic naturally occurring accent differences, which typically include multiple sounds. Single words (in shifted and unshifted pronunciations) and nonwords were recorded to assess perceptual learning using a lexical decision task (Maye et al., 2008; Weatherholtz, 2015). We chose to compare the Pleasant Shifted and Unpleasant Shifted conditions with a single pleasant control condition, as opposed to pleasant and unpleasant control conditions because we had no *a priori* reason to believe that listening to an unpleasant voice in the control condition would have changed response patterns compared to a pleasant voice control.

### 2.1. Materials

A trained phonetician with acting training and experience read the six-minute passage of Pinocchio from Weatherholtz (2015) in three guises: Control, Shifted, and Unpleasant Shifted. The Control guise was her normal reading voice (story duration = 4 minutes, 58 seconds). The Shifted guise was her normal voice, but she also pronounced the monophthongal back vowels according to the lowering chain shift exploited in Weatherholtz's design (story duration = 5 minutes, 24 seconds). The Unpleasant Shifted guise also included the back vowel lowering chain shift, in addition to being monotonous and creaky (story duration = 6 minutes, 30 seconds). Also recorded were the items of the lexical decision task. These items included the filler words, nonwords, and shifted back vowel items from Weatherholtz (2015). These single word and nonword items were produced in the both the speaker's pleasant normal voice and the unpleasant voice guise. Recordings were made using a head-mounted AKG C250 microphone connected to a SoundDevices 2.0 USB Preamp in a sound-attenuated cubicle and digitized at 44.1 kHz.

These produced stories naturally varied in duration with the Unpleasant Shifted guise being the slowest. Mean word duration in the Unpleasant Shifted guise was 323 ms, whereas it was 263 ms in the Shifted guise and 249 ms in the Control guise. These duration differences were present in the vowels, and **Table 1** provides duration summary statistics for monophthongal primary and secondary stressed vowels for the three guises. An ANOVA with duration as the dependent measure and Condition and Front/Back Vowel as independent measures was run. There was a significant effect of Condition [ $F(2, 1452) = 121.12, p < 0.001$ ] and

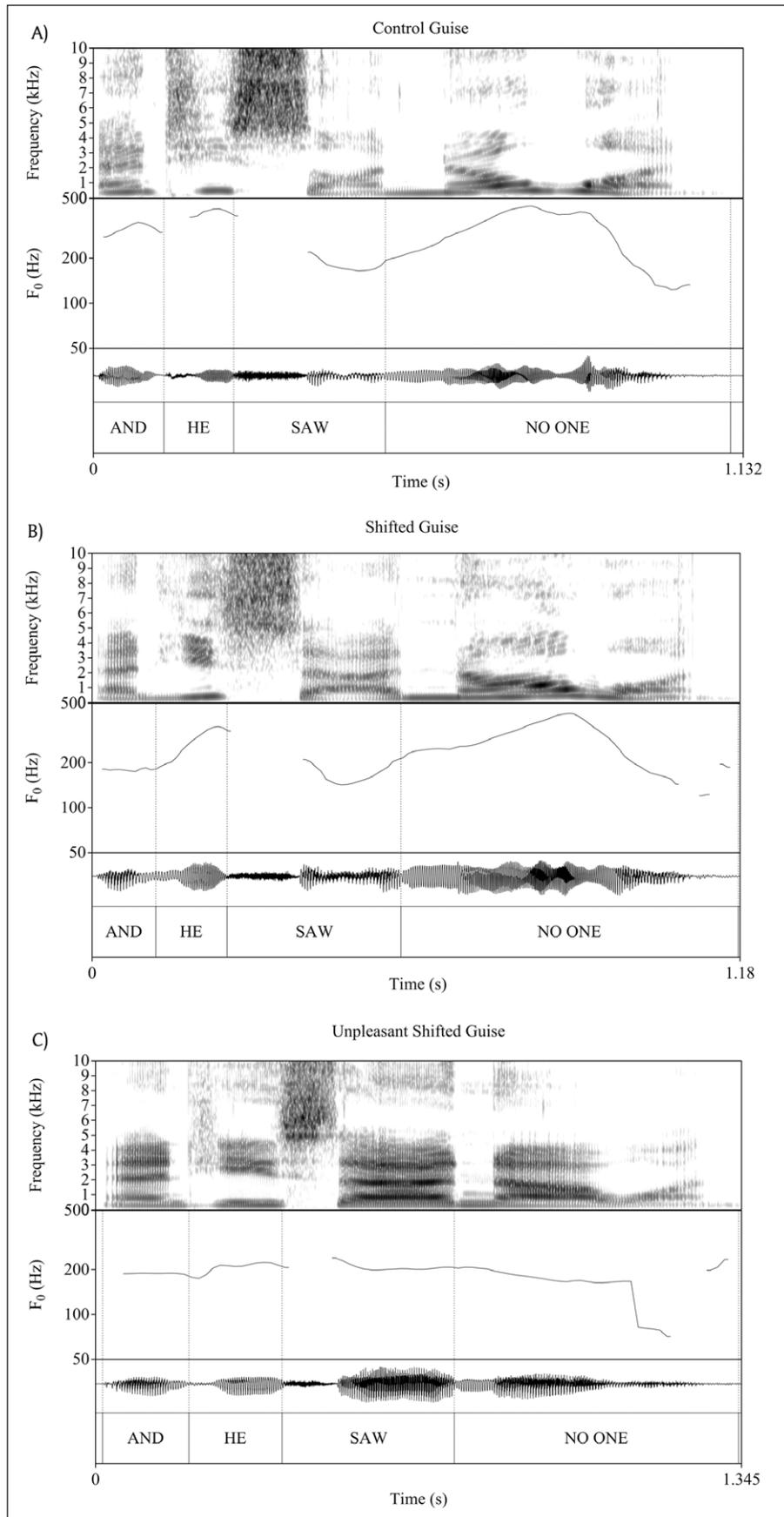
**Table 1:** Means and standard deviations in milliseconds for the monophthongal vowels in the three guises.

Vowel (in Arpabet)	Vowel (in IPA)	Standard	Shifted	Unpleasant Shifted
IY	i	97 (35)	107 (36)	154 (46)
IH	ɪ	70 (31)	67 (30)	87 (32)
EY	e	114 (27)	123 (42)	162 (44)
EH	ɛ	92 (34)	96 (36)	134 (42)
AE	æ	134 (45)	135 (41)	186 (44)
AA	ɑ	110 (41)	138 (50)	176 (71)
OW	o	120 (60)	145 (70)	184 (83)
UH	ʌ	80 (29)	149 (49)	187 (50)
UW	u	123 (53)	125 (52)	153 (49)

Front/Back Vowel [ $F(1, 1452) = 101.19, p < 0.001$ ], as well as an interaction between the two [ $F(2, 1452) = 6.56, p = 0.0012$ ]. T-tests confirmed that the back vowels in the Unpleasant Shifted guise ( $M = 178$  ms,  $SD = 73$ ) were longer in duration than those in the Shifted ( $M = 141$  ms,  $SD = 62$ ) guise [ $t(234) = 13.7, p < 0.001$ ] and longer than those in the Control ( $M = 114$  ms,  $SD = 54$ ) guise [ $t(234) = 19.8, p < 0.001$ ]. The vowels in the Normal Shifted guise were also longer than those in the Control guise [ $t(234) = 10.84, p < 0.001$ ]. T-tests also confirmed that the slower speech rate extended to the realization of the front vowels in Unpleasant Shifted guise ( $M = 143$  ms,  $SD = 52$ ) compared to the Shifted ( $M = 104$  ms,  $SD = 43$ ) guise [ $t(250) = 18.02, p < 0.001$ ] and the Control ( $M = 98$  ms,  $SD = 40$  ms) guise [ $t(250) = 13.67, p < 0.001$ ]. For the front vowels, the Shifted guise and the Control guise did not differ significantly from one another [ $t(250) = 1.45, p = 0.15$ ]. **Figure 1A–C** illustrates a sample utterance from each guise, illustrating the clear prosodic differences in both duration and  $f_0$  contour between the Unpleasant Shifted guise and the Control and Shifted guises.

## 2.2. Confirmatory vowel analyses

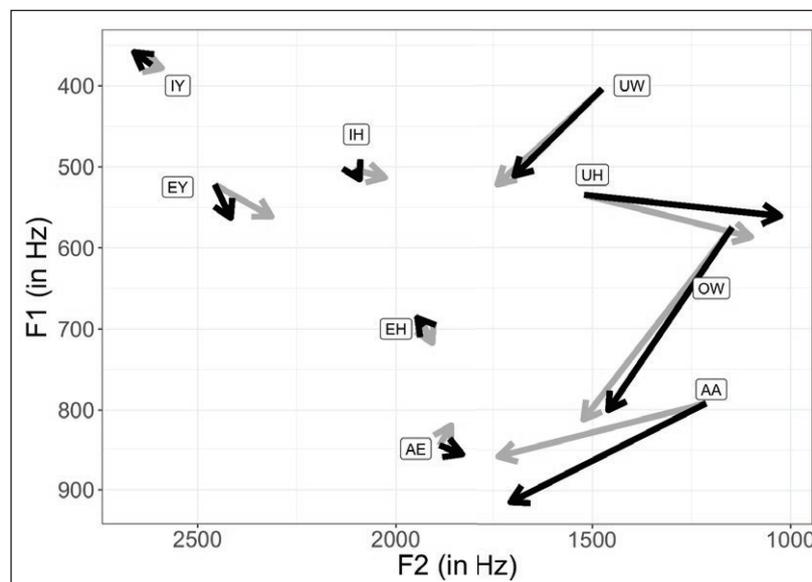
The story guises were force aligned using the FAVE Forced Aligner (Rosenfelder et al., 2014) and vowel boundaries were hand-corrected. **Table 2** provides counts for how many times each back vowel is realized in a stressed word in the story. The local accent has merged items in the lexical sets THOUGHT and LOT (e.g., the caught/cot merger), and this category is represented as /ɑ/ in IPA and AA in Arpabet in **Table 2** and **Figure 2**. Given that all of our stimuli were naturally produced, there is natural variability in the realization of all of the items. To confirm, however, that overall the vowel shifts in the two shifted story guises are equivalent, F1 and F2 values at the midpoint were estimated for monophthongs in stressed positions for the three guises. These values were estimated using the LPC function in Praat, and extracted in a supervised fashion where the number of LPC coefficients and the frequency range over which the formants were estimated were hand-adjusted as necessary on a by-token basis to ensure accurate estimation. These midpoint F1 and F2 values are illustrated in **Figure 2**. The black arrows point to the F1 by F2 mean of the vowel in the Shifted guise and the lighter gray arrows point to the F1 by F2 mean of the vowel in the Unpleasant Shifted guise. The two shifted guises are roughly equivalent, although the Unpleasant Shifted guise shows more evidence of reduction (i.e., less extreme F1 and F2 values) for the shifted vowels, which is in line with the intentionally less expressive storytelling of the unpleasant guise.



**Figure 1:** Spectrogram, waveform, textgrid, and pitch tracks of the utterance “and he saw no one” from each of the three guises: Control (top panel), Shifted (middle panel), Unpleasant Shifted (bottom panel).

**Table 2:** Counts for stressed tokens of the back vowels in the Pinocchio exposure story.

Vowel (in Arpabet)	Vowel (in IPA)	Count
AA	ɑ	44
AE	æ	42
EH	ɛ	86
EY	e	40
IH	ɪ	45
IY	i	38
OW	o	129
UH	ʌ	27
UW	u	35



**Figure 2:** Vowel F1 and F2 midpoints for the back vowels and the front monophthongal vowels for reference. The plot characters are in Arpabet. The black arrows point to the F1 by F2 mean of the vowel in the Shifted guise and the lighter gray arrows point to the F1 by F2 mean of the vowel in the Unpleasant Shifted guise.

To assess these patterns statistically, we conducted a series of ANOVAs. We first compared the F1 values by Vowel and Condition for the unshifted front vowels (IY, IH, EH, EY, AE), which confirmed an effect of Vowel [ $F(4, 738) = 492, p < 0.001$ ]. The unshifted front vowels were also assessed in terms of F2. There was the anticipated effect of vowel [ $F(4, 738) = 333.49, p < 0.001$ ] and an effect of Condition [ $F(2, 738) = 4.58, p = 0.01$ ]. A Tukey test exploring this effect of Condition found that none of the comparisons were reliably different from one another, establishing minimal differences in the front vowel space across the guises. To confirm differences in the back vowel space, we conducted identical analyses for the back vowels (UW, UH, OW, AA). The F1 analysis revealed an effect of Vowel [ $F(3, 693) = 220.6, p < 0.001$ ], Condition [ $F(2, 693) = 140.1, p < 0.001$ ], and an interaction between the two [ $F(6, 693) = 10.5, p < 0.001$ ]. The F2 analysis for the back vowels similarly revealed main effects of Vowel [ $F(3, 693) = 68.71, p < 0.001$ ] and Condition [ $F(2, 693) = 105.14, p < 0.001$ ], in addition to their interaction [ $F(6, 693) = 35.52, p < 0.001$ ]. Follow-up ANOVAs were conducted for each vowel to confirm

the effect of condition. For UW, the F1 and F2 analyses revealed effects of Condition [F1:  $F(2, 102) = 18.29, p < 0.001$ ; F2:  $F(2, 102) = 14.52, p < 0.001$ ]. Tukey tests comparing the difference conditions confirmed that along the F1 and F2 dimensions, UW in the Shifted conditions was higher than Standard guise ( $p < 0.001$  for both F1 and F2 for the two shifted guises), and that the two Unpleasant and Pleasant Shifted conditions did not differ from each other. For the UH vowel, there was no significant deviation along the F1 dimension between conditions [ $F(2, 78) = 3.12, p = 0.05$ ], but there was along the F2 dimension [ $F(2, 78) = 27.4, p < 0.001$ ]. The two Shifted guises were different from the Standard guise ( $p < 0.001$  in Tukey test), but not from one another. For the OW vowel, there was an effect of Condition for both the F1 and F2 analyses [F1:  $F(2, 384) = 120.7, p < 0.001$ ; F2:  $F(2, 384) = 89.43, p < 0.001$ ]. Tukey tests confirmed that for both F1 and F2 dimensions, the back vowel Shifted guises differed from the Standard guise ( $p < 0.001$  for both comparisons), but did not differ from each other at the 0.05 level. For the vowel AA, the ANOVAs revealed effects of condition for both F1 [ $F(2, 129) = 10.42, p < 0.001$ ] and F2 [ $F(2, 129) = 92.18, p < 0.001$ ]. Tukey tests revealed robust differences in F2, with both the shifted guises differing from the standard guise ( $p < 0.001$  for both shifted guises), and not from one another. In terms of F1, the normal Shifted guise was reliably different from the Standard guise at the 0.001 level but the Unpleasant Shifted guise was different at the 0.04 level.

Taken together, the analyses of vowel midpoints establish that the Shifted guises differed from the Standard guise along the back vowel dimensions, and not in terms of the front vowels for the means of the vowel targets. We also assessed whether the guises were matched in terms of the category variability, as more variable categories may be more challenging to adapt to. We estimated category variability using a measure of category dispersion. For each vowel for each guise, the Euclidean distance of a token's F1/F2 distance was calculated to that vowel's mean F1/F2 for each guise. The summary statistics for these calculations are given in **Table 3**. To quantify these patterns statistically, the Euclidean distance of each token was entered as the dependent measure in an ANOVA with Condition and Vowel as the independent variables, which found main effects of Vowel [ $F(8, 1431) = 13.38, p < 0.001$ ], Condition [ $F(2, 1431) = 15, p < 0.001$ ], and the interaction between Vowel and Condition [ $F(16, 1431) = 3, p < 0.001$ ]. To unpack the Vowel by Condition interaction, this ANOVA was followed up by a series of ANOVAs for each vowel to determine the effect of Condition. Starting with the critical back vowels, for AA and UW there was no effect of Condition on vowel category dispersion. The

**Table 3:** Means and standard deviations of category variability for each vowel in each guise, given in Euclidean distance in Hertz.

Vowel (in Arpabet)	Vowel (in IPA)	Standard	Shifted	Unpleasant Shifted
IY	ɪ	161 (163)	156 (140)	137 (106)
IH	ɪ	284 (159)	258 (134)	209 (130)
EY	E	119 (66)	133 (75)	110 (55)
EH	ɛ	219 (122)	215 (181)	148 (86)
AE	Æ	234 (92)	210 (161)	152 (90)
AA	ɑ	176 (121)	216 (161)	168 (147)
OW	O	202 (123)	278 (163)	231 (132)
UH	ʌ	268 (188)	139 (198)	117 (210)
UW	u	228 (150)	208 (104)	166 (89)

analysis for UH established an effect of Condition [ $F(2, 78) = 4.6, p = 0.013$ ]. Tukey tests determined that the Unpleasant Shifted guise was less variable than the Standard guise ( $p = 0.018$ ). The category dispersion analysis for OW found an effect of Condition [ $F(2, 384) = 9.78, p < 0.001$ ]. Tukey tests confirmed that the shifted OW vowel for the normal Shifted guise was more variable than that of the Unpleasant Shifted guise ( $p = 0.018$ ) and the Standard guise ( $p < 0.001$ ). For the front vowels, there were no significant vowel category dispersion differences between the guises for EY, IH, and IY. For AE, there was an effect of Condition [ $F(2, 123) = 5.3, p = 0.006$ ]. A Tukey test found that the Unpleasant Shifted guise was significantly less variable than the Standard guise ( $p = 0.005$ ). The analysis for EH also found an effect of Condition [ $F(2, 255) = 7.4, p < 0.001$ ], with a Tukey test confirming that the Unpleasant Shifted condition was less variable than both the normal Shifted guise ( $p = 0.004$ ) and the Standard guise ( $p = 0.002$ ).

In sum, for the critical back vowels on which listeners must adapt, the normal shifted guise was more variable in terms of OW compared to the Unpleasant Shifted and Standard control guises, and the Unpleasant Shifted guise was less variable with UH compared to the Standard control guise. Given this, it seems that the category variability of these guises do not differ substantially and reliably across vowels in a way that would impact retuning.

### 3. Voice preferences

While our voice actor crafted the Unpleasant guise to be unpleasant, it is necessary to establish that listeners find the guise socially dispreferred. We establish these preferences in this first experiment.

Naturally occurring accents elicit judgments of language attitudes related to the social stereotypes around these accents (Zahn & Hooper, 1985). Research on language attitudes has identified status and solidarity as the two dimensions on which listeners base their social evaluations. Status is operationalized as a dimension related to competence, intelligence, and socio-economic standing. Solidarity, on the other hand, is seen more as a measure of ingroup membership, with accents that are rated higher in solidarity being judged as more friendly and sociable (Dragojevic & Giles, 2016). We use the status and solidarity dimensions of the language attitude literature to establish whether the three social guises we designed elicited the intended social preferences. We quantified social preference in a forced choice task where listeners evaluated utterances matched for lexical content in terms of social status and solidarity in a pairwise fashion. We predict that listeners will rate the Control and Shifted guises as higher in status and solidarity than the Unpleasant Shifted guise, and that listeners will rate the Control guise as higher than the Shifted guise in these social dimensions in utterances that include shifted back vowels.

#### 3.1. Materials

The three story guises were separated into shorter, interpretable utterances based on naturally occurring prosodic boundaries and breath groups. This resulted in a total of 108 utterances. Thirteen of these utterances did not contain any back vowels. The remaining 95 utterances had anywhere from one to five back vowels. The number of utterances which contained four or five back vowels totaled 13, so these were used with the 13 utterances without back vowels in the voice preferences task. Thus, a total of 26 utterances were used from each of the three guises.

#### 3.2. Participants

Eighty-six participants completed the judgment task. Thirty-nine of these participants were nonnative speakers of English, leaving 47 native English speakers for the analysis (33 female, 13 male, 1 nonbinary, mean age = 21). Native speaker was defined as being fluent in English and having learnt the language before the age of five (based on self-report

in a language background survey). All the participants were undergraduates recruited from the University of British Columbia and compensated with partial course credit.

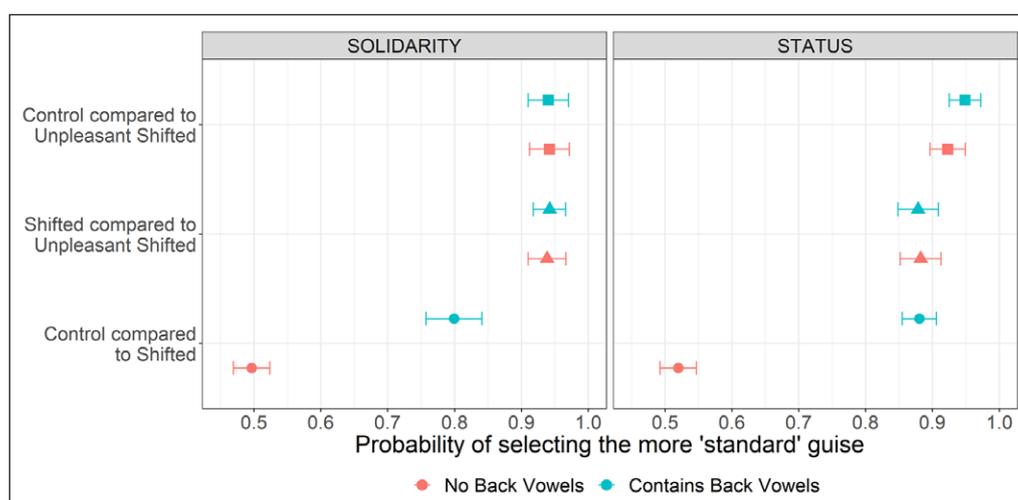
### 3.3. Procedure

Participants were run up to four at a time at individual workstations in sound-attenuated cubicles and outfitted with AKG K240 Studio Headphones and a serial response box. The experiment was presented in E-Prime 2.0 (Psychology Software Tools, Inc., 2012). Participants were presented with two utterances matched in lexical content and asked to evaluate which voice was higher in status or solidarity in two separate counter-balanced blocks. In the status block, listeners were told to determine which voice was higher in status, which was defined as sounding more intelligent and competent. In the solidarity block, listeners determined which voice was higher in solidarity, which was defined for them as sounding more friendly and sociable. Each utterance from each guise was presented with each matched utterance from the other two guises, such that each listener made a total of 78 comparisons for each social dimension. The order of the guises for a given utterance was counterbalanced across participants. Utterance order was randomized individually for each participant.

### 3.4. Results

The data were coded according to the probability that listeners selected the guise that was predicted to be socially preferred. That is, selecting to the Control guise when it was paired with the Shifted guise or the Unpleasant Shifted was scored as 1, and selecting the Shifted guise when it was paired with the Unpleasant Shifted guise was scored as 1. Listeners were predicted to always disprefer the Unpleasant Shifted guise due to its general style and voice quality, regardless of whether the utterance contained a shifted back vowel. When comparing the Control and Shifted guises, listeners should not show a preference for items without back vowels, as there should be no audible phonetic differences between these utterances, but listeners should prefer the Control guise to the Shifted guise for utterances which include back vowels.

**Figure 3** shows listeners' responses for each guise pairing for utterances with and without low back vowels for the two social dimensions. Listeners by and large dispreferred the



**Figure 3:** Listener responses to the Voice Preferences task described in Experiment 1. Responses were coded in terms of listeners' likelihood of choosing the more comparatively standard guise (Control > Shifted > Unpleasant Shifted). Responses are separated by Social evaluation dimension, and according to whether or not the utterance included back vowels. Error bars represent by-listener standard error.

Unpleasant Shifted guise over the Shifted and Control guises, as illustrated by the strong probability of choosing the Shifted and Control guises as higher in solidarity and social status. However, the solidarity evaluation was generally more aligned with the predicted social preferences. The comparison between the Control and Shifted guises is crucially dependent on whether the utterance contained a low back vowel, as there should be no difference between these guises in the absence of a low back vowel. As **Figure 3** illustrates, listeners' preferences for the Control or Shifted guise in the absence of a low back vowel cross 0.5 along both dimensions, and listeners have a preference for the Control guise over the Shifted guise in the low back vowel utterances, although this preference is not as robust as when a comparison involves the Unpleasant condition.

Listeners' pairwise preferences along the Solidarity and Status dimensions were analyzed in a logistic mixed effects model using the R (R Core Team, 2018) package *lme4* (Bates, Maechler, Bolker, & Walker, 2015). Button ordering and social dimension counterbalancing differences were collapsed. Guise comparison was contrast coded as follows: to compare the dispreference of the Shifted Unpleasant guise, Control-Shifted guise comparison = 2/3, the Control-Shifted Unpleasant guise comparison = -1/3, and the Shifted-Shifted Unpleasant guise comparison = -1/3; for the comparison of the whether the Shifted Unpleasant guise was dispreferred more strongly compared to the Shifted guise or the Control guise, Control-Shifted guise comparison = 0, Shifted-Shifted Unpleasant guise comparison = 1/2, and Control-Shifted Unpleasant guise comparison = -1/2; back vowel status was contrast coded (no Back Vowel = 1, Back Vowel = -1); likewise, Social Attribute was also contrast coded (Solidarity = 1, Status = -1). Listeners and Item were included as random intercepts. By-listener random slopes for guise comparison, back vowel, and social attribute were included, and by-item random slopes for guise comparison and social attribute were as well.<sup>1</sup>

The significant intercept establishes that listeners were more likely to select the voice that was designed to be more socially preferred. The model results are presented in **Table 4**, which confirm the patterns observed in **Figure 3**. Crucially, the interaction of Control/Shifted Comparison versus Control/Shifted Unpleasant and Shifted/Shifted Unpleasant with Vowel illustrates that listeners differed in their social evaluation of the Control-Shifted Comparison based on whether there was a back vowel in the utterance. Ratings were significantly higher for the Solidarity dimension, but the attribute did not interact with the voice guises.

### 3.5. Discussion

Listeners socially evaluated the voices in matched utterances in terms of social solidarity and social status in a pairwise fashion. While the analysis revealed small differences in listeners' evaluations of these guises in terms of solidarity and status, a couple of crucial conclusions that gloss over these small differences can be made. Listeners robustly rank the Unpleasant Shifted voice as lower in social status than the Control and Shifted guises with utterances both lacking and including the shifted back vowels. This indicates that listeners have negative associations with the Unpleasant voice guise. When comparing the Control and Shifted guises in utterances without back vowels, listeners hover around 50% in terms of which guise is preferred; this indicates that the speaker produced a uniformly pleasant voice guise in the Control reading and in the Shifted reading. In utterances with

<sup>1</sup> The model used in this analysis was  $ACC \sim \text{Guise Comparison.Helmert.Coded} * \text{Vowel} * \text{Attribute} + (1 + \text{Guise Comparison.Helmert.Coded} * \text{Vowel} * \text{Attribute} | \text{Subject}) + (1 + \text{Guise Comparison.Helmert.Coded} * \text{Attribute} | \text{Chunk})$ .

**Table 4:** Model results for the Voice Preferences task.

	Estimate	Standard Error	z value	Pr(> z )
Intercept	4.02	0.37	10.94	<0.001***
Control/Shifted Comparison versus Control/Shifted Unpleasant and Shifted/Shifted Unpleasant	-3.8	0.55	-6.85	<0.001***
Control/Shifted Unpleasant versus Shifted/Shifted Unpleasant	-2.29	0.85	-2.7	0.007**
No Back Vowel versus Back Vowel	-0.6	0.21	-2.92	0.003**
Solidarity versus Status	0.71	0.35	2	0.04*
Control/Shifted Comparison versus Control/Shifted Unpleasant and Shifted/Shifted Unpleasant: No Back Vowel versus Back Vowel	-1.27	0.36	-3.54	<0.001***
Control/Shifted Unpleasant versus Shifted/Shifted Unpleasant: No Back Vowel versus Back Vowel	0.55	0.59	0.94	0.35
Control/Shifted Comparison versus Control/Shifted Unpleasant and Shifted/Shifted Unpleasant: Solidarity versus Status	-0.96	0.52	-1.84	0.07
Control/Shifted Unpleasant versus Shifted/Shifted Unpleasant: Solidarity versus Status	0.11	0.82	0.13	0.9
Back Vowel versus No Back Vowel: Solidarity versus Status	0.08	0.21	0.39	0.69
Control/Shifted Comparison versus Control/Shifted Unpleasant and Shifted/Shifted Unpleasant: Back Vowel versus No Back Vowel: Solidarity versus Status	-0.3	0.35	-0.86	0.39
Control/Shifted Unpleasant versus Shifted/Shifted Unpleasant: Back Vowel versus No Back Vowel: Solidarity versus Status	-0.28	0.58	-0.47	0.64

back vowels, where the Control and Shifted conditions differ, listeners preferred the guise that produced the local standard and not the shifted back vowel pronunciations.

In short, this experiment demonstrates that listeners have robust and uniform social evaluations of these voices, clearly preferring the more standard voice quality and prosody of the Control and Shifted guises over that of the Unpleasant guise. Given the forced choice nature of our task, we are unable to state whether this is a large difference, only that it is a robust difference. Listeners rate the Control guise as having higher social status than the Shifted guise in trials that illustrate the shifted back vowels. Having confirmed these robust social preferences, we now examine whether listeners perceptually adapt more to the socially preferred guise.

#### 4. Perceptual learning with a lexical decision post-test

Following Maye et al. (2008) and Weatherholtz (2015), we use a lexical decision task to assess whether listeners perceptually adapt to the novel accent with the shifted back vowels. If through exposure to a story with the shifted back vowels listeners adapt their lexical templates, listeners in the Shifted and Unpleasant Shifted conditions should identify items with shifted back vowels that were also in the story (e.g., items like *wooden* pronounced as [wodŋ]) as words at higher rates than listeners in the control condition.

If listeners exposed to shifted back vowels also more broadly adjust their phonological expectations, they should also identify novel shifted back vowel words that were not in the story as words at higher rates than listeners in the control condition. These results would be a replication of Weatherholtz (2015). Crucially, given the social dimensions of these guises, if listeners attend less to socially dispreferred voices, we anticipate that listeners exposed to the Unpleasant Shifted guise should learn less than those exposed to the Shifted guise.

#### 4.1. Materials

Materials for this experiment included the full versions of the stories read in the Control, Shifted, and Unpleasant Shifted guises. Following the lexical decision blocks in Weatherholtz (2015), there were 60 nonwords, 100 filler words, 20 trained shifted back vowel items that were words included in the story, and 40 novel shifted back vowel items. The lexical decision items were presented with speech styles that were matched for the story, but were separate recordings made as single words. That is, for the Control and Shifted guises, listeners continued to hear the speaker's normal pleasant voice, while the items in the Unpleasant Shifted condition continued to be presented in the less pleasant voice style. Mean F1 and F2 values for the vowels for the monophthongal words in the lexical decision task are shown in **Table 5**. No statistical analysis was attempted given the small number of items for each vowel and stimuli type in the test items. Following Weatherholtz, no back vowel items with the low back vowel (/ɑ/, which would shift to /æ/) are included in the test to avoid a merger in the test phase. Four pseudorandomized stimuli lists were designed, which ensured that no two back vowel items occurred sequentially or within the first two trials, following Weatherholtz (2015). Listeners logged their responses by pressing assigned buttons on a serial response box.

#### 4.2. Participants

A total of 151 participants completed this experiment, though the participants who did not complete the language background survey or were not native speakers of English were not included in the analysis, leaving 98 participants (68 female, 28 male, 1 nonbinary; mean age = 22). Thirty-three of these individuals were assigned to the Control condition, 31 to the Shifted condition, and 33 to the Unpleasant Shifted condition. All the participants were undergraduates recruited from the University of British Columbia and were compensated with partial course credit.

**Table 5:** Vowel category means in Hertz for the test stimuli in the lexical decision task.

Vowel (Arpabet)	Vowel (IPA)	Standard		Unpleasant standard		Shifted		Unpleasant shifted	
		F1	F2	F1	F2	F1	F2	F1	F2
IY	i	312	2862	364	2703	–	–	–	–
IH	ɪ	494	2286	579	2089	–	–	–	–
EY	e	473	2701	505	2398	–	–	–	–
EH	ɛ	738	2009	709	1913	–	–	–	–
AE	æ	907	1977	851	1922	–	–	–	–
OW	o	510	960	570	1078	846	1275	872	1345
UH	ʌ	461	1205	617	1556	548	939	488	999
UW	u	342	1028	426	1183	448	1796	445	1816

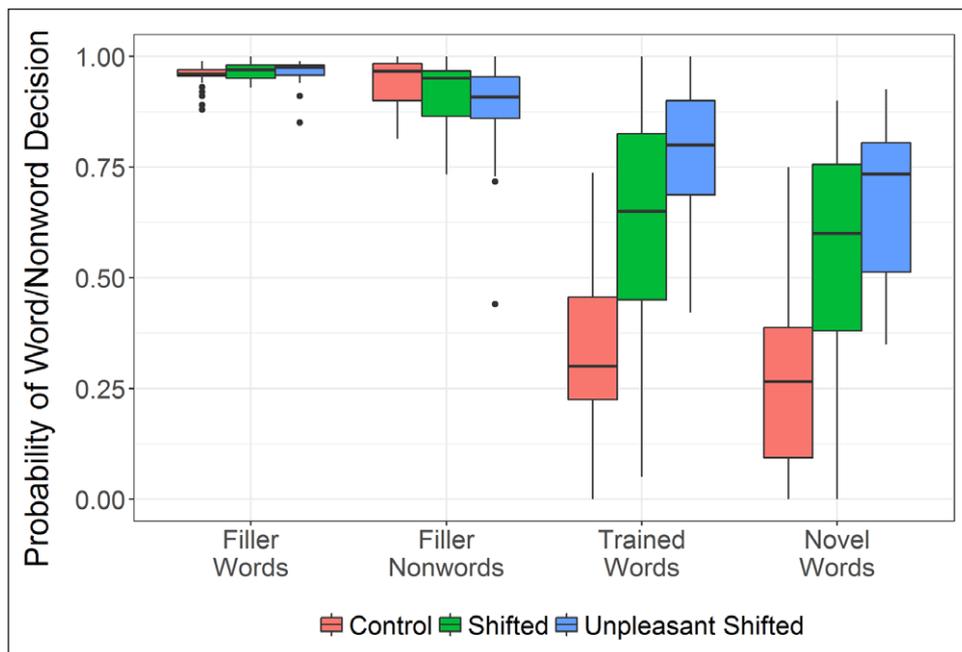
### 4.3. Procedure

Participants were run up to four at a time at individual workstations in sound-attenuated cubicles and outfitted with AKG K240 Studio Headphones and a serial response box. The experiment was presented in E-Prime 2.0 and composed of two parts. In the first part, participants listened to one of the story guises over headphones, which was accompanied by a static image of the marionette Pinocchio on the computer monitor. Listeners were asked to listen quietly. In the second part, listeners completed a lexical decision task. They were presented with the lexical decision items over the headphones and asked to determine whether each item was a word in English or not a word in English. Participants responded using the assigned buttons on the button box (e.g., 1 = word, 5 = nonword) and were asked to respond as quickly and accurately as possible. Responses were allowed for up to three seconds; if participants did not respond during this time, the experiment progressed to the next trial. Participants were given self-paced breaks every 50 items during the lexical decision task. Lastly, they completed a language background survey. The experiment took around half an hour to complete.

### 4.4. Results

Trials with no responses were removed from the data set (98 trials, less than 0.05% of the data set). Responses were coded as correct (1) for word responses to filler words and trained or novel words with shifted back vowels and for nonword responses to the nonword items. **Figure 4** presents these results as a boxplot with the mean accuracy for each trial type for each listener. Listener behaviour was consistent for responses to the filler word and nonword items, but considerably more variable for the critical items with shifted vowels across all of the conditions.

To facilitate interpretation, four logistic mixed effects models with accuracy as the dependent variable were fit for each level of the lexical status variable (filler word, nonword, trained back vowel word, novel back vowel word). Condition was contrast coded as follows: for the Control versus Shifted Conditions comparison, Control = 2/3,



**Figure 4:** Listener accuracy on word/nonword decisions for the four item types in Experiment 2. Responses to Trained and Novel shifted back vowel words were coded as correct when a word response was given.

Shifted =  $-1/3$ , Unpleasant Shifted =  $-1/3$ ; for the Shifted versus Unpleasant Shifted comparison, Control = 0, Shifted =  $1/2$ , Shifted Unpleasant =  $-1/2$ . This allows us to assess whether the two shifted conditions differ from the control condition and whether the shifted conditions differ from each other. Listener was included as a random intercept. Item was entered as a random intercept with condition as a by-item random slope.<sup>2</sup>

The model outputs are presented in **Table 6**. Condition assignment did not affect listeners' accurate endorsement of filler words, but Condition did affect performance with other item types. For nonword fillers, listeners in the shifted conditions were significantly less accurate than listeners in the control condition. For both trained and novel shifted back vowels, listeners in the shifted conditions were more likely to call these items words than listeners in the control condition. Additionally, for both the shifted conditions, listeners in the Unpleasant Shifted condition were more likely to endorse these items as words than listeners exposed to the shifted pronunciations in the pleasant Shifted condition. These patterns are visualized in **Figure 4**.

#### 4.5. Discussion

The results of this task clearly indicate that an unpleasant voice does not impede lexical retuning as measured by word endorsement rates of shifted items. Listeners exposed to the novel pronunciations through the Shifted or Unpleasant Shifted guise endorsed these novel pronunciations as words more than listeners who had not been exposed to these items in the Control condition. Listeners in the Unpleasant Condition endorsed the shifted back vowel pronunciations at higher rates and generalized the pronunciation pattern to novel unheard items more robustly than those who heard the same shifted items in a more socially preferred voice. It is possible that hearing an unpleasant voice changes the

**Table 6:** Model results for the Lexical Decision Task.

		Estimate	Standard Error	z value	Pr(> z )
<b>Filler words</b>	Intercept	5.01	0.26	19.02	<0.001***
	Control versus Shifted Conditions	-0.24	0.33	-0.73	0.47
	Shifted Unpleasant versus Shifted Pleasant	0.04	0.45	0.09	0.93
<b>Nonwords</b>	Intercept	3.08	0.18	17.4	<0.001***
	Control versus Shifted Conditions	0.67	0.29	2.33	0.02
	Shifted Unpleasant versus Shifted Pleasant	0.31	0.32	0.97	0.33
<b>Trained Shifted Back Vowels</b>	Intercept	0.66	0.42	1.56	0.12
	Control versus Shifted Conditions	-2.88	0.42	-6.86	<0.001***
	Shifted Unpleasant versus Shifted Pleasant	-1.53	0.58	-2.65	0.008**
<b>Novel Shifted Back Vowels</b>	Intercept	0.04	0.3	0.14	0.89
	Control versus Shifted Conditions	-2.45	0.40	-6.17	<0.001***
	Shifted Unpleasant versus Shifted Pleasant	-1.46	0.50	-2.93	0.003**

<sup>2</sup> The following code was used for the analysis in Experiment 2: `glmer(ACC ~ Condition.Helmert.Coded + (1 | Subject) + (1 + Condition.Helmert.Coded | Word)` for each of the four lexical levels.

threshold for accepting novel pronunciations at test; however, given our decision to not include an unpleasant control condition, we are unable to vet this interpretation.

Listeners in the shifted conditions were less accurate on nonword identification than those in the control condition. This could be due to a broader relaxing of criteria for lexical templates generally, but the magnitude of the loosening with respect to nonwords is smaller than the increase in word endorsement for the shifted words. Importantly, this pattern indicates we cannot make specific claims about the direction of perceptual learning, as this could simply be evidence of a relaxation of lexical template matching thresholds. Nevertheless, what is crucial here is that despite robust social preferences for the Shifted guise, listeners in the Unpleasant Shifted guise learned to accept novel pronunciations more than those who are exposed to the Shifted guise.

## 5. General discussion

The goal of this study was to determine whether social dispreference leads to an attenuation of perceptual learning. Experiment 1 used evaluative social dimensions from the language attitude literature to establish that the vocal guise which was designed to sound socially unpleasant or dispreferred was indeed robustly identified by listeners as being lower in status and solidarity than the control guise and the shifted guise. Listeners selected the control and shifted guise as having higher social status and higher solidarity than the unpleasant shifted guise 94% and 91% of the time, respectively. The shifted guise was also judged as being lower in status and solidarity than the control voice only on trials which contained shifted back vowels, on which trials listeners selected the control guise as socially preferred on 84% of the trials, compared to 51% of the trials without back vowels. These results indicate that listeners indeed robustly dispreferred the unpleasant guise which was monotone and had creaky voice quality, regardless of whether it displayed an unfamiliar shifted back vowel accent. Did these social evaluations affect perceptual learning?

There is evidence that listeners spontaneously phonetically imitate voices or accents they find socially preferable (Babel, 2010, 2012; Yu, Abrego-Collier, & Sonderegger, 2013), and while these effects could be the result of an attentional spotlight that privileges the subphonetic detail in perceptual processing or lexical encoding (Sumner et al., 2014), socially-guided phonetic accommodation could also be a wholly production-based implicit decision process. Using evidence from a selective adaptation paradigm, Zheng and Samuel (2017) recently argued that effects of increased perceived nonnative accent associated with Asian faces (Rubin, 1992; Yi, Phelps, Smiljanic, & Chandrasekaran, 2013; Babel & Russell, 2015) are the result of a post-perceptual interpretation or decision process and *not* the result of the actual perception of accented speech. Perceptual learning, like selective adaptation, has been argued to be the result of veridical changes to perception and not post-perceptual decision biases (Clarke-Davidson, Luce, & Sawusch, 2008). Thus, perceptual retuning offers a nice test case to assess whether social dispreference attenuates perceptual learning, as would be the case if listeners attended less to the phonetic detail of socially dispreferred voices or accents. Counter to such predictions, the results of Experiment 2 indicate that listeners adapt to the novel pronunciations they hear during a story regardless of whether the vocal guise is socially dispreferred or not. Listeners in the lexical decision task who heard the novel pronunciations in *The Adventures of Pinocchio* in the Shifted and Unpleasant Shifted voice identified words heard in the context of the story that contained the back vowel chain shift and novel words that were not included in the story as real words at much higher rates than listeners who had been exposed to the standard pronunciations in the control condition. These results replicate the findings

of Maye et al. (2008) and Weatherholtz (2015). These results also demonstrate that negative social evaluation does not negatively impact perceptual adaptation. Listeners expressed a clear dispreference for the voice guise we designed to be unpleasant, but despite these robust social preferences, these do not seem to guide listeners' retuning of lexical templates. While listeners in the control condition tended not to endorse the items with shifted back vowels as words in the lexical decision task, those who were exposed to either the Unpleasant Shifted or the Pleasant Shifted guise did. And, in fact, listeners in the unpleasant shifted guise seem to have generalized their acceptance to novel back vowel items more than listeners who were exposed to the more pleasant shifted guise. It appears to be the case that despite being socially dispreferred, the unpleasant voice elicited more perceptual adjustments. It is possible that while the social dispreference was robust, it was not a large enough social dispreference to sway any lexical retuning mechanisms. The forced choice nature of our social evaluation task prevents us from making claims about the size of the social preference.

Clarke-Davidson et al. (2008) suggest that perceptual adaptation is a phonetic retuning effect and not a decision bias. It may be the case that social evaluations, whether they be positive or negative, are post-perceptual decisions or interpretations (to use the term of Zheng & Samuel, 2017), and thus, when negative, have no inhibitory effect on perceptual adaptation. Novel pronunciations and speech styles may rather elicit attention by virtue of their novelty or atypicality, eliciting robust adaptive responses, as has been shown in phonetic accommodation (Babel, McGuire, Walters, & Nicholls, 2014). In fact, this heightened adaptation to novel pronunciations and unique voices is exactly what would be predicted by episodic models of spoken word processing where low familiarity voices show less competition and thus exert more of an influence on the perceptual system (Goldinger, 1998). Another possible explanation for the increased generalization in the lexical decision task for listeners exposed to the Unpleasant Shifted guise is that this guise had a slower speech rate, and the vowels in this guise were the longest in duration. Listeners would have been presented with longer exemplars of the back vowel shift, which may have facilitated adaptation. While the longer durations certainly may have boosted lexical retuning, such an interpretation does not counter the basic finding of these experiments: Listeners preferred the Shifted guise to the Unpleasant guise, but learned from the Unpleasant guise in spite of these social preferences. It is possible that despite the social dispreference, listeners learned from the Unpleasant guise *because* of the slower speech rate, which may have drawn attention to the vowels or increased the salience of the shift. On the other hand, the slow speech rate likely contributed to the dispreference of the guise, making it impossible to disentangle these interpretations with these materials.

The mechanisms underlying adaptation to small, targeted phonetic shifts like the back vowel shift modeled in this set of experiments and other experiments within the lexically guided perceptual learning canon may well be different than the more global adjustments that are necessary for naturally occurring dialect and non-native accent differences. Importantly, like Maye et al. (2008), our study only used one speaker, and does not examine generalization to multiple speakers, as would be the case with perceptual learning of an entire dialect. While the back vowel shift modeled here might be an acceptable proxy for subtle regional accent differences within a larger speech community, it certainly lacks the robust and multidimensional shifts of larger dialect differences. Such larger dialect differences introduce a more challenging task for the listener, and there is neurolinguistic evidence that listeners recruit attentional resources to decipher the more challenging signal-to-phonological mapping for such dialect differences (Van Engen & Peelle, 2014; Yi, Smiljanic, & Chandrasekaran, 2014). This recruitment of executive resources like attention

are likely a function of listener effort, which may relate to social motivations and preferences in comprehending unfamiliar highly dissimilar (compared to one's own) accents.

Together, these results invite us to speculate that social preferences may exert themselves as more of an influence on implicit choices in production, as opposed to highlighting phonetic detail in perceptual processing. For example, consider the case of the children acquiring gender-specific phonetic patterns prior to the onset of the physiological changes which would underlie sex or gender-based phonetic differences (e.g., Sachs, Lieberman, & Erickson, 1973) or children adopting local dialect patterns in lieu of the dialect of their caregivers (e.g., Chambers, 1992; Payne, 1980). The current results cast doubt on these being cases of socially selective perceptual learning. Rather, it appears more likely that these may be social influences on production. Again, however, we reiterate the challenge of disentangling real world social preferences from real world familiarity and experiences in voices.

While these results are far from conclusive they are a crucial first step in our attempts to query the role of social preferences in perceptual learning. Using adaptation-like paradigms is important to establish whether social influences on behavioural results are a reflection of perceptual warping or post-perceptual interpretations. Establishing that social effects in spoken language recognition stem from post-perceptual decision weighting mechanisms, for example, make them no less interesting, but simply facilitate our understanding of where within the complicated sensory and cognitive system a particular aspect of communication and language use lies.

## Acknowledgements

Thanks to Karina Wong, Stephanie Chung, and Jennifer Abel for their indispensable contributions to this project. Thank you to Kodi Weatherholtz for sharing his original stimuli with us and engaging in many productive conversations about his dissertation work. This work was supported at various points by SSHRC Grant 435-2014-1673 to Eric Vatikiotis-Bateson and SSHRC Grant 435-2017-0136 to Molly Babel.

## Competing Interests

The authors have no competing interests to declare.

## References

- Babel, M. 2010. Dialect divergence and convergence in New Zealand English. *Language in Society*, 39(4), 437–456. DOI: <https://doi.org/10.1017/S0047404510000400>
- Babel, M. 2012. Evidence for phonetic and social selectivity in spontaneous phonetic imitation. *Journal of Phonetics*, 40(1), 177–189. DOI: <https://doi.org/10.1016/j.wocn.2011.09.001>
- Babel, M., & McGuire, G. 2015. Perceptual fluency and judgments of vocal aesthetics and stereotypicality. *Cognitive Science*, 39(4), 766–787. DOI: <https://doi.org/10.1111/cogs.12179>
- Babel, M., McGuire, G., Walters, S., & Nicholls, A. 2014. Novelty and social preference in phonetic accommodation. *Laboratory Phonology*, 5(1), 123–150. DOI: <https://doi.org/10.1515/lp-2014-0006>
- Babel, M., & Russell, J. 2015. Expectations and speech intelligibility. *The Journal of the Acoustical Society of America*, 137(5), 2823–2833. DOI: <https://doi.org/10.1121/1.4919317>
- Bates, D., Maechler, M., Bolker, B., & Walker, S. 2015. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1), 1–48. DOI: <https://doi.org/10.18637/jss.v067.i01>

- Bradlow, A. R., & Alexander, J. A. 2007. Semantic and phonetic enhancements for speech-in-noise recognition by native and non-native listeners. *The Journal of the Acoustical Society of America*, 121(4), 2339–2349. DOI: <https://doi.org/10.1121/1.2642103>
- Bradlow, A. R., & Bent, T. 2008. Perceptual adaptation to non-native speech. *Cognition*, 106(2), 707–729. DOI: <https://doi.org/10.1016/j.cognition.2007.04.005>
- Chambers, J. K. 1992. Dialect acquisition. *Language*, 673–705. DOI: <https://doi.org/10.2307/416850>
- Clarke, C. M., & Garrett, M. F. 2004. Rapid adaptation to foreign-accented English. *The Journal of the Acoustical Society of America*, 116(6), 3647–3658. DOI: <https://doi.org/10.1121/1.1815131>
- Clarke-Davidson, C. M., Luce, P. A., & Sawusch, J. R. 2008. Does perceptual learning in speech reflect changes in phonetic category representation or decision bias? *Attention, Perception, & Psychophysics*, 70(4), 604–618. DOI: <https://doi.org/10.3758/PP.70.4.604>
- Clopper, C. G., Tamati, T. N., & Pierrehumbert, J. B. 2016. Variation in the strength of lexical encoding across dialects. *Journal of Phonetics*, 58, 87–103. DOI: <https://doi.org/10.1016/j.wocn.2016.06.002>
- Cutler, A., Mehler, J., Norris, D., & Segui, J. 1987. Phoneme identification and the lexicon. *Cognitive Psychology*, 19(2), 141–177. DOI: [https://doi.org/10.1016/0010-0285\(87\)90010-7](https://doi.org/10.1016/0010-0285(87)90010-7)
- Dragojevic, M., & Giles, H. 2016. I don't like you because you're hard to understand: The role of processing fluency in the language attitudes process. *Human Communication Research*, 42(3), 396–420. DOI: <https://doi.org/10.1111/hcre.12079>
- Eisner, F., & McQueen, J. M. 2005. The specificity of perceptual learning in speech processing. *Perception & Psychophysics*, 67(2), 224–238. DOI: <https://doi.org/10.3758/BF03206487>
- Francis, A. L., MacPherson, M. K., Chandrasekaran, B., & Alvar, A. M. 2016. Autonomic nervous system responses during perception of masked speech may reflect constructs other than subjective listening effort. *Frontiers in Psychology*, 7, 263. DOI: <https://doi.org/10.3389/fpsyg.2016.00263>
- Francis, A. L., & Nusbaum, H. C. 2002. Selective attention and the acquisition of new phonetic categories. *Journal of Experimental Psychology: Human Perception and Performance*, 28(2), 349–366. DOI: <https://doi.org/10.1037/0096-1523.28.2.349>
- Goggin, J. P., Thompson, C. P., Strube, G., & Simental, L. R. 1991. The role of language familiarity in voice identification. *Memory & Cognition*, 19(5), 448–458. DOI: <https://doi.org/10.3758/BF03199567>
- Goldinger, S. D. 1998. Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, 105(2), 251–279. DOI: <https://doi.org/10.1037/0033-295X.105.2.251>
- Goldstone, R. L. 1998. Perceptual learning. *Annual Review of Psychology*, 49(1), 585–612. DOI: <https://doi.org/10.1146/annurev.psych.49.1.585>
- Hay, J., Warren, P., & Drager, K. 2006. Factors influencing speech perception in the context of a merger-in-progress. *Journal of Phonetics*, 34(4), 458–484. DOI: <https://doi.org/10.1016/j.wocn.2005.10.001>
- Holt, R. F., & Bent, T. 2017. Children's use of semantic context in perception of foreign-accented speech. *Journal of Speech, Language, and Hearing Research*, 60(1), 223–230. DOI: [https://doi.org/10.1044/2016\\_JSLHR-H-16-0014](https://doi.org/10.1044/2016_JSLHR-H-16-0014)
- Jesse, A., & McQueen, J. M. 2011. Positional effects in the lexical retuning of speech perception. *Psychonomic Bulletin & Review*, 18(5), 943–950. DOI: <https://doi.org/10.3758/s13423-011-0129-2>

- Johnson, K. 1997. Speech perception without speaker normalization: An exemplar model. In: Johnson, K., & Mullennix, J. W. (eds.), *Talker Variability in Speech Processing*, 145–165. San Diego: Academic Press.
- Johnson, K., Strand, E. A., & D’Imperio, M. 1999. Auditory–visual integration of talker gender in vowel perception. *Journal of Phonetics*, 27(4), 359–384. DOI: <https://doi.org/10.1006/jpho.1999.0100>
- Kleinschmidt, D. F., & Jaeger, T. F. 2015. Robust speech perception: Recognize the familiar, generalize to the similar, and adapt to the novel. *Psychological Review*, 122(2), 148–203. DOI: <https://doi.org/10.1037/a0038695>
- Kraljic, T., & Samuel, A. G. 2005. Perceptual learning for speech: Is there a return to normal? *Cognitive Psychology*, 51(2), 141–178. DOI: <https://doi.org/10.1016/j.cogpsych.2005.05.001>
- Kraljic, T., & Samuel, A. G. 2006. Generalization in perceptual learning for speech. *Psychonomic Bulletin & Review*, 13(2), 262–268. DOI: <https://doi.org/10.3758/BF03193841>
- Kraljic, T., & Samuel, A. G. 2007. Perceptual adjustments to multiple speakers. *Journal of Memory and Language*, 56(1), 1–15. DOI: <https://doi.org/10.1016/j.jml.2006.07.010>
- Lattner, S., Maess, B., Wang, Y., Schauer, M., Alter, K., & Friederici, A. D. 2003. Dissociation of human and computer voices in the brain: Evidence for a preattentive gestalt-like perception. *Human Brain Mapping*, 20(1), 13–21. DOI: <https://doi.org/10.1002/hbm.10118>
- Lippi-Green, R. 1997. *English with an accent: Language, ideology, and discrimination in the United States*. Psychology Press.
- Maye, J., Aslin, R. N., & Tanenhaus, M. K. 2008. The weckud wetch of the wast: Lexical adaptation to a novel accent. *Cognitive Science*, 32(3), 543–562. DOI: <https://doi.org/10.1080/03640210802035357>
- McAuliffe, M., & Babel, M. 2016. Stimulus-directed attention attenuates lexically-guided perceptual learning. *The Journal of the Acoustical Society of America*, 140(3), 1727–1738. DOI: <https://doi.org/10.1121/1.4962529>
- Munson, B., & Babel, M. under review. The Phonetics of Sex and Gender. In: Katz, W., & Assmann, P. (eds.), *Routledge Handbook of Phonetics*.
- Norris, D., McQueen, J. M., & Cutler, A. 2003. Perceptual learning in speech. *Cognitive Psychology*, 47(2), 204–238. DOI: [https://doi.org/10.1016/S0010-0285\(03\)00006-9](https://doi.org/10.1016/S0010-0285(03)00006-9)
- Payne, A. 1980. Factors controlling the acquisition of the Philadelphia dialect by out-of-state children. *Locating language in time and space*, 1, 143–78.
- Perrachione, T. K., & Wong, P. C. 2007. Learning to recognize speakers of a non-native language: Implications for the functional organization of human auditory cortex. *Neuropsychologia*, 45(8), 1899–1910. DOI: <https://doi.org/10.1016/j.neuropsychologia.2006.11.015>
- Psychology Software Tools, Inc. 2012. E-Prime 2.0 [computer software].
- R Core Team. 2018. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria. <<https://www.R-project.org>>.
- Reby, D., McComb, K., Cargnelutti, B., Darwin, C., Fitch, W. T., & Clutton-Brock, T. 2005. Red deer stags use formants as assessment cues during intrasexual agonistic interactions. *Proceedings of the Royal Society B*, 272, 941–947. DOI: <https://doi.org/10.1098/rspb.2004.2954>
- Reinisch, E., & Holt, L. L. 2014. Lexically guided phonetic retuning of foreign-accented speech and its generalization. *Journal of Experimental Psychology: Human Perception and Performance*, 40(2), 539. DOI: <https://doi.org/10.1037/a0034409>

- Reinisch, E., Weber, A., & Mitterer, H. 2013. Listeners retune phoneme categories across languages. *Journal of Experimental Psychology: Human Perception and Performance*, 39(1), 75–86. DOI: <https://doi.org/10.1037/a0027979>
- Rendall, D., Owren, M., Weerts, E., & Hienz, R. 2004. Sex differences in the acoustic structure of vowel-like vocalizations in baboons and their perceptual discrimination by baboon listeners. *Journal of the Acoustical Society of America*, 115, 411–421. DOI: <https://doi.org/10.1121/1.1635838>
- Rosenfelder, I., Fruehwald, J., Evanini, K., Seyfarth, S., Gorman, K., Prichard, H., & Yuan, J. 2014. FAVE (Forced Alignment and Vowel Extraction) Program Suite v1.2.2. DOI: <https://doi.org/10.5281/zenodo.22281>
- Rubin, D. L. 1992. Nonlanguage factors affecting undergraduates' judgments of nonnative English-speaking teaching assistants. *Research in Higher Education*, 33(4), 511–531. DOI: <https://doi.org/10.1007/BF00973770>
- Sachs, J., Lieberman, P., & Erickson, D. 1973. Anatomical and cultural determinants of male and female speech. *Language Attitudes: Current trends and prospects*, 74–84.
- Scharenborg, O., & Janse, E. 2013. Comparing lexically guided perceptual learning in younger and older listeners. *Attention, Perception, & Psychophysics*, 75(3), 525–536. DOI: <https://doi.org/10.3758/s13414-013-0422-4>
- Smither, J. A.-A. 1993. Short term memory demands in processing synthetic speech by old and young adults. *Behavior and Information Technology*, 12, 330–335. DOI: <https://doi.org/10.1080/01449299308924397>
- Sumner, M., Kim, S. K., King, E., & McGowan, K. B. 2014. The socially weighted encoding of spoken words: A dual-route approach to speech perception. *Frontiers in Psychology*, 4(1015), 1–13. DOI: <https://doi.org/10.3389/fpsyg.2013.01015>
- Thompson, C. P. 1987. A language effect in voice identification. *Applied Cognitive Psychology*, 1(2), 121–131. DOI: <https://doi.org/10.1002/acp.2350010205>
- Todd, R. M., Talmi, D., Schmitz, T. W., Susskind, J., & Anderson, A. K. 2012. Psychophysical and neural evidence for emotion-enhanced perceptual vividness. *Journal of Neuroscience*, 32(33), 11201–11212. DOI: <https://doi.org/10.1523/JNEUROSCI.0155-12.2012>
- Van Engen, K. J., & Peelle, J. E. 2014. Listening effort and accented speech. *Frontiers in Human Neuroscience*, 8, 577. DOI: <https://doi.org/10.3389/fnhum.2014.00577>
- Weatherholtz, K. 2015. *Perceptual learning of systemic cross-category vowel variation* (Unpublished doctoral dissertation). The Ohio State University, Columbus, OH.
- Werker, J. F., & Tees, R. C. 1984. Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant behavior and development*, 7(1), 49–63. DOI: [https://doi.org/10.1016/S0163-6383\(84\)80022-3](https://doi.org/10.1016/S0163-6383(84)80022-3)
- White, M., Rajkumar, R., Ito, K., & Speer, S. R. 2009. Eye tracking for the online evaluation of prosody in speech synthesis: Not so fast!. *Proceedings of INTERSPEECH*, 2523–2526.
- Winters, S. J., Levi, S. V., & Pisoni, D. B. 2008. Identification and discrimination of bilingual talkers across languages. *The Journal of the Acoustical Society of America*, 123(6), 4524–4538. DOI: <https://doi.org/10.1121/1.2913046>
- Yi, H. G., Phelps, J. E., Smiljanic, R., & Chandrasekaran, B. 2013. Reduced efficiency of audiovisual integration for nonnative speech. *The Journal of the Acoustical Society of America*, 134(5), EL387–EL393. DOI: <https://doi.org/10.1121/1.4822320>
- Yi, H. G., Smiljanic, R., & Chandrasekaran, B. 2014. The neural processing of foreign-accented speech and its relationship to listener bias. *Frontiers in Human Neuroscience*, 8. DOI: <https://doi.org/10.3389/fnhum.2014.00768>
- Yu, A. C. L., Abrego-Collier, C., & Sonderegger, M. 2013. Phonetic imitation from an individual-difference perspective: Subjective attitude, personality and “autistic” traits. *PloS one*, 8(9), e74746. DOI: <https://doi.org/10.1371/journal.pone.0074746>

- Zahn, C. J., & Hopper, R. 1985. Measuring language attitudes: The speech evaluation instrument. *Journal of Language and Social Psychology*, 4(2), 113–123. DOI: <https://doi.org/10.1177/0261927X8500400203>
- Zeelenberg, R., Wagenmakers, E.-J., & Shiffrin, R. M. 2004. Nonword Repetition Priming in Lexical Decision Reverses as a Function of Study Task and Speed Stress. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30(1), 270–277. DOI: <https://doi.org/10.1037/0278-7393.30.1.270>
- Zhang, X., & Samuel, A. G. 2014. Perceptual learning of speech under optimal and adverse conditions. *Journal of Experimental Psychology: Human Perception and Performance*, 40(1), 200–217. DOI: <https://doi.org/10.1037/a0033182>
- Zheng, Y., & Samuel, A. G. 2017. Does seeing an Asian face make speech sound more accented? *Attention, Perception, & Psychophysics*, 1–19. DOI: <https://doi.org/10.3758/s13414-017-1329-2>

**How to cite this article:** Babel, M., Senior, B. and Bishop, S. 2019 Do social preferences matter in lexical retuning? *Laboratory Phonology: Journal of the Association for Laboratory Phonology* 10(1): 4, pp. 1–22. DOI: <https://doi.org/10.5334/labphon.133>

**Submitted:** 02 January 2018    **Accepted:** 11 January 2019    **Published:** 25 February 2019

**Copyright:** © 2019 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

 *Laboratory Phonology: Journal of the Association for Laboratory Phonology* is a peer-reviewed open access journal published by Ubiquity Press.

**OPEN ACCESS** 