Exploring effects of brief daily exposure to unfamiliar accent on listening performance and cognitive load

Drew J. McLaughlin[1,2]*, Melissa M. Baese-Berk[3,4] and
Kristin J. Van Engen[1]

[1]Department of Psychological and Brain Sciences, Washington University in St. Louis, St. Louis, MO,
United States, [2]Basque Center on Cognition, Brain and Language, San Sebastian, Spain, [3]Department
of Linguistics, University of Oregon, Eugene, OR, United States, [4]Department of Linguistics, University
of Chicago, Chicago, IL, United States

Introduction: Listeners rapidly "tune" to unfamiliar accented speech, and some evidence also suggests that they may improve over multiple days of exposure. The present study aimed to measure accommodation of unfamiliar second language- (L2-) accented speech over a consecutive 5-day period using both a measure of listening performance (speech recognition accuracy) and a measure of cognitive load (a dual-task paradigm).

Methods: All subjects completed a dual-task paradigm with L1 and L2 accent on Days 1 and 5, and were given brief exposure to either L1 (control group) or unfamiliar L2 (training groups) accent on Days 2–4. One training group was exposed to the L2 accent via a standard speech transcription task while the other was exposed to the L2 accent via a transcription task that included implicit feedback (i.e., showing the correct answer after each trial).

Results: Although overall improvement in listening performance and reduction in cognitive load were observed from Days 1 to 5, our results indicated neither a larger benefit for the L2 accent training groups compared to the control group nor a difference based on the implicit feedback manipulation.

Discussion: We conclude that the L2 accent trainings implemented in the present study did not successfully promote long-term learning benefits of a statistically meaningful magnitude, presenting our findings as a methodologically informative starting point for future research on this topic.

Introduction

Evidence suggests that listeners can rapidly "tune" to unfamiliar accented speech, thereby improving their ability to understand a given speaker over time. For second language- (L2-) accented speech, improvements to listening performance (often measured with transcription/repetition accuracy, or "intelligibility") can be facilitated by exposure to a single accented speaker, to multiple speakers with the same accent, or even to a variety of speakers with different accents (Bradlow and Bent, 2008; Sidaras et al., 2009; Baese-Berk et al., 2013). Similarly, the cognitive demands of speech processing have been shown to rapidly decrease following exposure to L2-accented speech in single-session experiments (Clarke and Garrett, 2004; Brown et al., 2020). Based on correlational evidence, it also appears that the efficiency and accuracy of L2 accent processing depends on a listener's prior (real world) experience: More experienced listeners typically process L2 accent faster

and more accurately (Kennedy and Trofimovich, 2008; Porretta et al., 2020). However, empirical evidence connecting these two literatures is lacking, with few studies that have examined perceptual accommodation of L2 accent across multiple days (or weeks, etc.). In the present study, we take a first step toward filling this empirical gap. Across five consecutive daily sessions, we sought to document changes in listening performance and cognitive load for (previously unfamiliar) L2 accent.

Accent experience: a theoretical framework

Because spoken language varies from talker to talker, due to both idiosyncratic differences and accent, listeners have to be adaptable when mapping complex acoustic input onto linguistic representations in the mental lexicon (Bent and Baese-Berk, 2021). Changes to representations (and/or decision processes, see Xie et al., 2023[1]) based on listeners' global and recent exposure are supported by multiple leading language models, including exemplar (Johnson, 1997; Pierrehumbert, 2001), non-analytic episodes (Goldinger, 1998; Nygaard and Pisoni, 1998), and Bayesian inference (i.e., "the ideal adaptor"; Kleinschmidt and Jaeger, 2015) models. Under these frameworks, it is posited that listeners create categories systematically linking social groupings and phonetic patterns, including accent-specific representations. On this view, listeners' prior experience with a given accent ought to determine their ability to efficiently and accurately process speech produced with that accent. In the same way that processing speech produced by a familiar talker is faster (Newman and Evers, 2007; Magnuson et al., 2021), processing speech produced in a familiar accent ought to be faster.

Correlational evidence aligns with the supposition that the ability to process an L2 accent accurately and efficiently can be developed over time with sufficient (real world) exposure. For example, Kennedy and Trofimovich (2008) found that both semantically meaningful and semantically anomalous Mandarin Chinese-accented sentences were transcribed with higher accuracy by L1 English listeners who had greater experience with Mandarin Chinese accent. Psychophysiological evidence from pupillometry and eye-tracking has also indicated a benefit of experience: Task-evoked pupil response in Porretta and Tucker (2019) indicated that L1 English listeners who had more experience with Mandarin Chinese accent processed Mandarin Chinese-accented English words (presented in noise) more easily, and gaze behavior in Porretta et al. (2020) demonstrated that greater experience with Mandarin Chinese accent also resulted in faster speech processing for Mandarin Chinese-accented English. Further behavioral evidence from L1 Dutch listeners also suggests that activation of L2-accented words depends on listener experience with the target accent. Witteman et al. (2013) used a cross-modal lexical decision task in which each trial participants were

presented auditorily with a German-accented word or non-word in Dutch followed by an orthographic probe word or non-word in Dutch. The listeners' task was to make lexical decisions for the visual probe items. Results indicated that listeners with less experience with German accent were less primed by auditory items that were strongly accented than participants with greater experience.

Altogether, these studies suggest a critical role of prior experience in L2 accent processing, aligning with predictions from episodic models of speech processing. What remains to be empirically determined, however, is the amount and rate of exposure that is necessary to observe a benefit of prior experience. Under all of the theoretical frameworks mentioned above (exemplar, non-analytic episodes, and Bayesian inference), listeners with more experience with a given accent ought to be more adept at processing speech produced with that accent. The present study aims to test these theoretical models of speech processing.

Accent experience: empirical evidence

Only a small number of studies have investigated the benefits of prolonged exposure to L2 accent using causal experimental design. Within a single experimental session, rapid improvements to listening performance and reductions of listening effort are well-documented (Clarke and Garrett, 2004; Bradlow and Bent, 2008; Sidaras et al., 2009; Baese-Berk et al., 2013; Brown et al., 2020). Beyond a single experimental session, however, investigations of sustained benefits of L2 accent exposure are limited, and those that exist have produced mixed results.

Examining both transcription accuracy and comprehension of Korean-accented English materials, Lindemann et al. (2016) found a sustained benefit of L2 accent exposure at 1–2 days post-training. The authors presented L1 English listeners with either L2 accent exposure or an explicit linguistic training (i.e., teaching phonemic differences between Korean and English, etc.). Notable for the present study, participants in the L2 accent exposure group completed a speech transcription task in which the correct sentence was presented each trial after submitting the typed response. At test, both training groups had better sentence transcription accuracy than a control group; comprehension scores were the same across groups. Thus, the results of Lindemann et al. demonstrate that a (brief) training session with L2 accent can lead to improvements in perceptual accuracy that last into subsequent day(s).

Sustained adaptation to L2 accent over a half-day (12-h) period—as well as generalization—was also demonstrated in a sleep consolidation study conducted by Xie et al. (2017). L1 English listeners in the study were trained with word-length stimuli from a Mandarin Chinese-accented talker, focusing on a key accented phoneme (/d/). All subjects completed a test with a novel Mandarin Chinese-accented talker immediately after training as well as a second iteration of this test 12-h later, but for half of the subjects this 12-h period spanned the day (e.g., 8 a.m. to 8 p.m.) and for the other half is spanned the night (e.g., 8 p.m. to 8 a.m.). In both groups, retention of the training benefit was observed. Critically, however, the overnight group showed

---

1 Computational evidence from Xie et al. (2023) suggests that many of the benefits observed in adaptive speech perception experiments may be explained by multiple mechanisms, including: (1) changes to phonemic representations, (2) pre-linguistic signal normalization, and (3) changes in post-perceptual decision-making criteria.

unique generalization of learning to another Mandarin Chinese-accented speaker and phonemic category (/t/), suggesting that sleep consolidation promoted generalization of learning.

Whether benefits of L2 accent exposure are retained over intervals longer than 1 day, however, remains unclear. Bieber and Gordon-Salant (2021), for example, failed to find evidence of a training benefit in test sessions administered 1 week after training. In their study, the authors examined accent-generalizable learning (i.e., performance on a novel/untrained accent following training with multiple other accents) for speech presented in six-talker babble. They employed a dual-task paradigm (similar to the one used in the present study), which combines a speech transcription task with a simultaneous reaction time-based visual task (it is assumed that with finite cognitive resources, reaction times will slow for the secondary visual task as the demands of the primary speech task increase). L1 English-speaking young adults and older adults with and without hearing loss completed three experimental sessions across approximately 3 weeks, where the beginning portion of Week 2's session and Week 3's session each served as measures of retention. Results indicated that within an experimental session listeners rapidly improved, as in prior work (Bradlow and Bent, 2008; Baese-Berk et al., 2013). However, the benefit of each prior week's training session on transcription accuracy was not retained (i.e., the beginning of the Week 2 and 3 sessions did not demonstrate improvement). Reaction times from the secondary measure, on the other hand, were significantly improved for Week 3's session, in particular, indicating that the cognitive load associated with L2 accent processing may have been reduced.[2]

Focusing on both listening performance and attitudes toward L2 speakers, Derwing et al. (2002) implemented what appears to be the longest L2 accent training protocol to date, occurring over an 8-week period. The authors sought to train L1 English listeners to better understand L2 (specifically, Vietnamese) accent, comparing the effects of a training with explicit phonetic lectures and a training with cross-cultural awareness lectures. Unfortunately, results of the study indicated no significant benefits of either training for speech transcription or comprehension. Attitude questionnaires, however, did reveal that both training groups showed increased empathy toward immigrants, and participants given explicit phonetic training reported increased confidence in their ability to understand L2 accent.

Altogether, the current body of empirical evidence suggests that benefits of L2 accent training sessions may persist into subsequent days but diminish over longer (week-long) intervals. Additionally, benefits observed for cognitive load may diverge from those observed for listening performance (i.e., recognition/transcription accuracy). Based on these observations, in the present study we sought to examine the benefits of a training protocol administered over multiple consecutive days. From Pre-Test to Post-Test, we also incorporated a measure of cognitive load (similar to Bieber and Gordon-Salant, 2021) to determine whether different benefits

may be observed for measures of listening performance vs. cognitive load.

The present study

In the present study, we implemented a combination of dual-task paradigms and a speech transcription tasks over a 5-day period. On Days 1 and 5, participants completed a Pre-Test and Post-Test (dual-task paradigm), and on Days 2, 3, and 4 participants completed exposure-based training sessions (speech transcription). Participants were randomly assigned to one of three groups for the training days: Control (no exposure to L2 accent or feedback during training), Exposure (exposure to L2 accent but no feedback during training), and Feedback (exposure to L2 accent and feedback during training). All groups had the exact same Pre- and Post-Test with both L1- and L2-accented speech stimuli.

We predicted that response times to the secondary task in the dual-task paradigm (an index of cognitive load) would be shorter on Day 5 than Day 1 for all groups, indicating improvement on the task itself. Critically, we expected that this improvement would be greater for the Exposure and Feedback training groups—particularly in the L2-accented speech condition—than it would be for the Control group. Additionally, we predicted that the Feedback group would show greater reduction in cognitive load than the Exposure group, given that the feedback manipulation provided lexical context to guide perceptual adaptation.

For listening performance (speech recognition accuracy) in the Pre-Test and Post-Test data, we had similar predictions, although we also anticipated the possibility that subjects may demonstrate reduced cognitive load without gains in listening performance (as in Bieber and Gordon-Salant, 2021). We expected that speech recognition scores from the primary task would be larger on Day 5 than Day 1 for all groups (indicating improvement on the task itself), and that the Exposure and Feedback training groups would improve more than the Control group in the L2-accented speech condition, in particular. We also predicted that the Feedback group would show greater improvement than the Exposure group.

Lastly, we planned to examine listening performance (speech transcription[3] accuracy) data from the training sessions on Days 2, 3, and 4. We predicted that, if any differences existed, they would be as follows: Higher scores for the Feedback group than the Exposure group overall, and an interaction with days reflecting greater improvement for the Feedback group over time.

---

2 One limitation of this finding is that, without a control condition, the reductions in cognitive demands cannot be solely attributed to the training. It is possible that familiarization with the secondary task led to improved reaction times.

---

3 The difference in terminology for the Pre-Test and Post-Test sessions (speech *recognition*) vs. the training sessions (speech *transcription*) corresponds to the different task demands. In the dual-task paradigm, participants heard target sentences and then repeated them aloud (because the secondary task required use of their hands to make responses). In the training sessions, however, there was no secondary (dual) task, so participants listened to the target sentences and then typed what they heard into a response box. Both measures are used to index *listening performance*.

Methods

The current study was approved by Washington University's Institutional Review Board. Due to the COVID-19 pandemic, the final version of the study deviated substantially from the pre-registered version (full details can be found in the Supplementary material).

Participants

Young adult subjects (age mean = 19.5; age range = 18–30) were recruited from Washington University in St. Louis's Psychology Participants Pool. Inclusion criteria (set via demographic filters in SONA Systems) selected for L1 English speakers with normal hearing and vision (or corrected-to-normal vision). Additional criteria on the SONA listing indicated that subjects should not sign up for the study if they had extensive exposure to Mandarin Chinese (for example, they should not speak Mandarin Chinese, have studied Mandarin Chinese, or have parents or roommates who are fluent in Mandarin Chinese). Subjects who did not complete all 5 days of the study were excluded from analyses.

Due to COVID-19-related recruitment issues, we decided to combine a pilot version of the experiment with the main dataset to reach more desirable sample size ($N = 160$). We report full details regarding the minor differences between the pilot and primary subject groups below. In brief, the two differences were: (1) During the dual-task sessions for the primary group but not the pilot group, the practice session provided feedback instructing participants to "speed up" if they took longer than 3 s to respond; and, (2) Additional measures of cognitive ability (not analyzed in the present manuscript) were not collected from the pilot participants. Results of all analyses remained the same when accounting for time of participation (i.e., when including a two-level fixed effect denoting "pilot" vs. "main" experiment status); because time of participation did not improve model fits or impact the outcomes for the effects of interest, this factor was dropped from all models.

After combining the two datasets, the sample size by group was as follows: Control $n = 54$, Exposure $n = 53$, Feedback $n = 53$. In the pilot version of the experiment, a total of 43 subjects participated. Two subjects were excluded from this sample for failing to complete all days of the study, one for reporting exposure to Chinese, and three for having an average reaction time in the dual-task paradigm > 3,000 ms (the significance of this cut-off is discussed further in the Procedures section). After exclusions, 37 valid subjects remained (by group: Control $n = 11$, Exposure $n = 14$, Feedback $n = 12$). For the main experiment, 152 subjects participated in total. Twenty-nine of these subjects were excluded for one the following reasons: Failing to complete all 5 days of the study (24), self-reporting too much prior exposure to Mandarin Chinese (four), and, in one case, self-reporting a receptive and productive language disorder. After exclusions, 123 valid subjects remained (by group: Control $n = 43$, Exposure $n = 39$, Feedback $n = 41$).

We report information about participants' language experience by random assignment group in Table 1. All participants reported English as their primary language, and as a language learned from birth. All participants reported at least one additional language (which is to be expected given high school language requirements in the U.S.). As can be seen in Table 1, a fairly large proportion of the sample can be classified as simultaneous (21%) or early (8%) bilinguals; these trends are unsurprising when considering current estimates of bilingualism in the United States (~1 in 5 speak a language other than English at home; Dietrich and Hernandez, 2022). Including bilingual status as an effect in the response time and accuracy analyses did not change the results or improve model fits, and was thus not included as a factor in the final models.

Materials

Auditory stimuli

Semantically anomalous sentences from the Semantically Normal Sentence Test (SNST; Nye and Gaitenby, 1974) were adapted for use in the present experiment. The SNST includes items with four keywords each (defined as any adjectives, verbs, or nouns) such as "the *wrong shot led* the *farm.*" This sentence set was selected with the aim of examining perception of L2 accent in quiet listening conditions while preventing ceiling effects for transcription accuracy. The original SNST set contains 200 items, and for the present study we created an additional 110 items. This provided enough unique items to avoid repeating auditory stimuli at any point during the study (i.e., more than 297 unique items total).

Recordings of these sentences were created in a sound-reduction booth using MOTU UltraLite-mk3 Hybrid microphone hardware and Audacity (version 2.4.2) run on iMac (version 10.15.7). For the L2 accent condition, we selected Mandarin Chinese-accented English. Six young adult, female speakers were recorded reading all of the semantically-anomalous items. To select three speakers for the present study, we piloted the stimuli with 253 participants, for a total of ~10 transcriptions per item (i.e., each participant listened to only 90 items). Across all items and responses, transcription performance for three of the speakers was fairly well-matched and met the experiment needs. These speakers were estimated to be 51.8, 53.1, and 55.8% intelligible. All talkers were proficient English speakers who began learning English in China as children (at ages 11, 8, and 5, respectively) but had only been in the United States for ~1 year of graduate-level studies.

For the L1-accented condition, three female L1 speakers of English from the Midwestern United States were recorded. Given that timing of responses in the dual-task paradigm was of critical interest, we decided to match speaking rate across the L1 and L2 speakers. Toward this goal, the L1 speakers were instructed to produce items at a typical speed, a slightly slower than normal speed, and a slower than normal speed. Items were selected based on their total duration in order to match the speaking rate across the L2 and L1 speakers. The final average stimuli length for the L2 speakers was 1,790 ms, and the final average stimuli length for the L1 speakers was 1,784 ms.

TABLE 1 Summary of participants' language experiences.

	Count of participants	Count of languages	Proportion of participants with one or more L2(s) acquired from:		
			Birth	≤age 5	≤age 10
All groups	**160**	**2.66 (0.69)**	**0.21**	**0.29**	**0.54**
Control	54	2.65 (0.69)	0.22	0.31	0.57
Exposure	53	2.67 (0.71)	0.17	0.21	0.56
Feedback	53	2.66 (0.69)	0.26	0.36	0.48

Standard deviations shown in parentheses. The bold value indicate that its summary of the below rows.

Questionnaire

Participants completed a questionnaire on Days 2, 3, and 4 of the study that assessed motivation (composite of three questions), self-perceived performance, and effort. The questions for the motivation composite score included the following: (1) How motivated were you to perform well-during the listening task? (1 = very unmotivated, 7 = very motivated); (2) How much did you like performing trials in the listening task? (1 = strongly dislike, 7 = strongly like); (3) How much did you desire to challenge yourself during the listening task? (1 = not at all, 7 = very much). The self-perceived performance question ("Please rate your performance on the listening task") included a scale from "1 = absolute worst" to "7 = absolute best", and the effort question ("How effortful did you find the listening task?") included a scale from "1 = not at all effortful" to "7 = very effortful."

Procedures

Overview

Subjects completed five experimental sessions lasting 30 min each over the course of five consecutive days (typically Monday through Friday). On Days 1 and 5, the primary speech perception task involved a dual-task paradigm, and on Days 2, 3, and 4, the primary speech perception task involved self-paced speech transcription only. The speech perception tasks were administered on a 21.5 inch iMac (version 10.15.7, "Catalina") and programmed with SuperLab (Cedrus, version 5). Audio was presented via circumaural Beyerdynamic DT 100 headphones.

Additional measures occurred after the primary experimental tasks on specific days of the week as follows: On Day 1, participants completed a demographic and language background questionnaire; on Day 2, they completed the Trail-Making Task (Arbuthnott and Frank, 2000); on Day 3, they completed a Stroop task (MacLeod, 1991); on Day 4, they completed the Word Auditory Recognition and Recall Measure (WARRM; a measure of working memory capacity; Smith et al., 2016); and for Days 2, 3, and 4, they completed a questionnaire each day to assess their motivation, self-perceived performance, and effort (method and results reported in Supplementary material). Note that in the pilot version of the experiment, the Trail-Making, Stroop, and WARRM tasks were not included. We do not report on these individual difference measures in the present study.

Dual-task paradigm (Days 1 and 5)

The dual-task paradigm included a speech perception primary task and a non-linguistic visual categorization secondary task (used previously in Strand et al., 2018; Brown et al., 2020). Participants were instructed that they would be completing both tasks simultaneously, but that the speech perception task was the primary and more important task.

Each trial, subjects were presented with a single auditory sentence. Their goal was to repeat the sentence at the end of the trial as accurately as possible. At the onset of the soundfile, two empty squares appeared on the screen. After an interstimulus interval (ISI) of 600–800 ms (in 100 ms intervals), a number between 1 and 8 appeared in either the left or the right box. Using a button box, participants were instructed to make either a left response or a right response depending on the following: If an odd number appeared (1, 3, 5, or 7), they were supposed to press the button on the opposite side as the box on the screen; if, however, an even number appears (2, 4, 6, or 8) they were supposed to press the button on the same side as the box on the screen. For example, the correct response for a 1 appearing in the left box on the screen was pressing the right button, and the correct response for a 2 appearing in the left box on the screen was pressing the left button. Participants were instructed to respond as quickly as possible while prioritizing accuracy. The timing of the ISI ensured that the demands of the secondary task occurred approximately midway through the presentation of the target sentences for the primary task. Thus, trials in which the demands of the primary task were greater should result in longer response times to the secondary task.

For the primary task, participants repeated the target sentence aloud after both their key press was made for the secondary task and the auditory stimulus was completed for the primary task. Verbal responses were recorded and scored for accuracy offline. Between trials, an ISI of 5,000, 5,500, or 6,000 ms occurred before automatic presentation of the next trial.

The combination of items for the primary and secondary tasks was randomized across participants. For the secondary task, the occurrence of each number at each of the two locations occurred randomly. For the primary task, auditory files were presented in a random order within a list used for practice trials (12 total) and a list used for regular trials (78 total). An equal number of trials for each accent condition and each speaker were included. For the regular trials, this resulted in 39 trials per accent, and 13 trials per speaker. Four counterbalanced orders were used to rotate which target sentences appeared on Day 1 vs. Day 5, and whether these targets were presented in the L1 vs. the L2 accent condition on a given day.

During the practice trials, a researcher remained in the room to observe the participant and confirm they were making responses in the correct order (i.e., button press and then verbal repetition). After the pilot version of the experiment was complete, we decided to add feedback to the practice session. If subjects took longer than 3,000 ms to respond with a button press after presentation of the number target, "Too slow!" appeared onscreen. Data from practice trials was excluded from analyses.

A 72-trial block of the secondary task (i.e., presented in isolation) was completed after the critical dual-task session was complete. In this block, subjects were only presented with numbers to sort, and no auditory input. Pilot subjects were not presented with this block, and, thus, we do not report on this data in the present paper.

Speech transcription task (Days 2, 3, and 4)

A speech transcription task was administered on training days (Days 2, 3, and 4) instead of the dual-task paradigm. For the training sessions, subjects were randomly assigned to one of three conditions: Control, Exposure, and Feedback. Subjects assigned to the Control group heard only the three L1-accented talkers on training days, while those assigned to the Exposure and Feedback groups heard only the three L2-accented talkers on training days. The key difference between the Exposure and Feedback groups was that subjects in the Feedback group were shown the correct target sentence after submitting their transcription each trial (thus providing implicit feedback on performance).

Participants completed 39 trials each session (13 trials per talker) presented in a randomized order. None of the target sentences repeated across training sessions or overlapped with target sentences from the dual-task sessions. Transcriptions were completed with a keyboard and self-paced. Subjects were instructed to do their best to spell accurately. After entering their transcriptions, participants were shown either a series of eight hashtags (Control and Exposure groups) or the target sentence (Feedback group) for 5,000 ms. An inter-stimulus interval of 3,000 ms occurred before presentation of the next trial.

Analysis

Model specifications: recognition accuracy data

Generalized linear mixed-effects regression was used to model the recognition accuracy data in R (version 4.0.4; R Core Team, 2021) with the *glmer()* function from the *lme4* package (Bates et al., 2015). Likelihood ratio tests were conducted to determine the significance of effects of interest, and *p*-values for model parameters were estimated using the *lmerTest* package (Kuznetsova et al., 2017). Recognition accuracy was treated as a grouped binomial, meaning that models predicted performance using two columns of data (number of correct words, number of incorrect/missed words) for each sentence. A logit link function was specified. Fixed effects included: Condition (dummy-coded levels: L1 accent, L2 accent), Session (dummy-coded levels: Pre-Test, Post-Test), Group (dummy-coded levels: Control, Exposure, Feedback), as well as
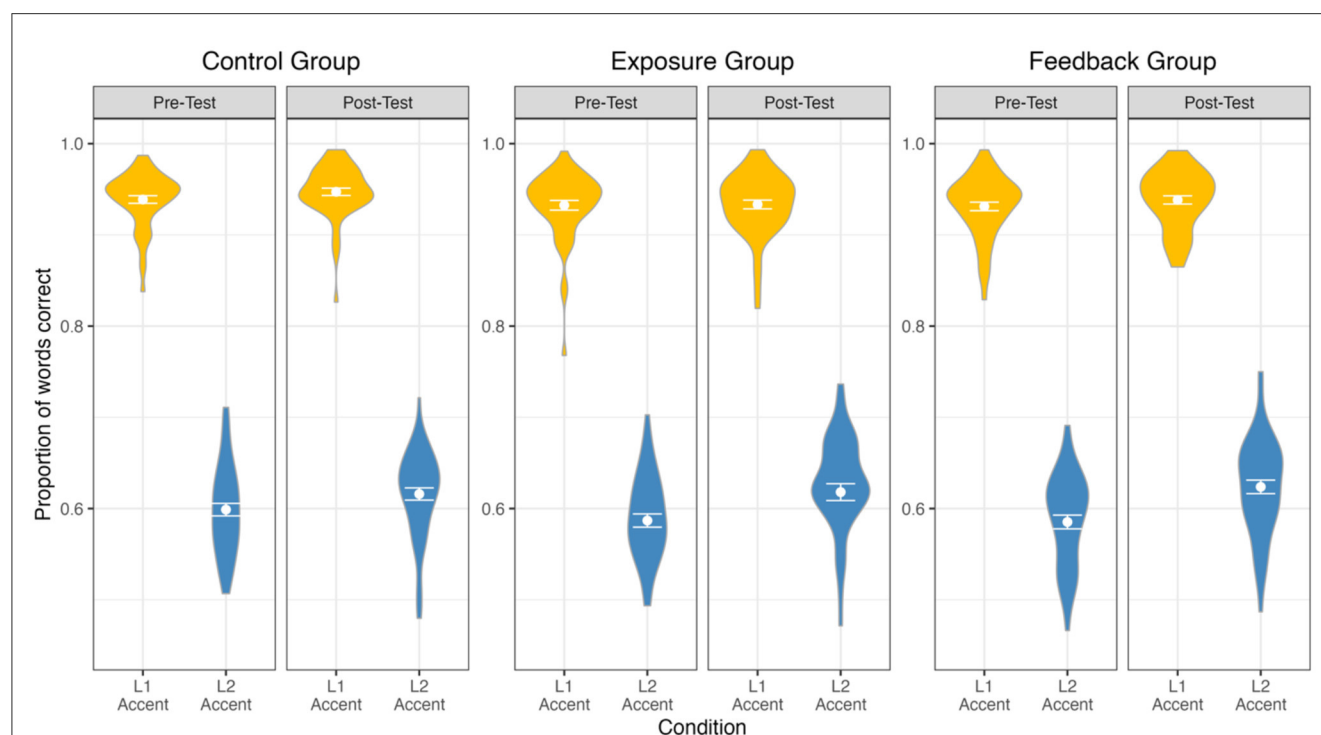

FIGURE 1
Recognition accuracy data from the dual-task paradigm at Pre-Test and Post-Test, for each group and accent condition, is presented with violin density distributions, mean points, and standard error bars.

TABLE 2 Log-likelihood model comparisons from analyses of dual-task recognition accuracy data.

Effect	χ^2	df	p
Condition	16,709	1	<0.001
Session	51.91	1	<0.001
Group	1.20	2	0.55
Condition: session	1.52	1	0.22
Condition: group	8.52	2	0.01
Session: group	4.02	2	0.13
Condition: session: group	4.56	2	0.10

df, degrees of freedom.

all possible two- and three-way interactions between Condition, Session, and Group. Random intercepts were included by item and by subject. Random slopes of Day and Condition were attempted but ultimately removed from all models due to issues with model singularity. Model syntax is provided in Supplemental materials.

Model specifications: response time data

For the response time data, linear mixed-effects regression was implemented with the *lmer()* function. Fixed effects included: Condition (dummy-coded levels: L1 accent, L2 accent), Session (dummy-coded levels: Pre-Test, Post-Test), Group (dummy-coded levels: Control, Exposure, Feedback), and all two- and three-way interactions between Condition, Session, and Group. In all models, random effects included random intercepts by subject and by item, and random slopes of Condition and Session by subject. Model syntax is provided in Supplemental materials.

Results

Pre–Test and Post–Test (dual–task paradigm) data

Recognition accuracy data from speech perception task

Recognition accuracy data from the dual-task paradigm is presented in Figure 1. We report all log-likelihood model comparisons in Table 2 and provide full model summaries in Supplemental materials. In brief, results indicated improved accuracy from Pre-Test to Post-Test ($\beta = 0.13$, $p < 0.001$), but this improvement was similar for all participant groups (non-significant three-way interaction of Condition, Session, and Group: $\chi^2 = 4.56$, $p = 0.10$).

As expected, overall performance for the L2 accent was significantly poorer than performance for the L1 accent ($\beta = -2.44$). The fixed effect of Group indicated that all groups had similar recognition accuracy, overall. Of the two-way interactions, only the interaction of Condition and Group significantly improved model fit. Model estimates indicated an overall smaller difference in performance between the L1 and L2 accent conditions for the Exposure group ($\beta = 0.15$, $p = 0.006$) and the Feedback group

($\beta = 0.12$, $p = 0.006$) compared to the Control group. As noted above, the log-likelihood model comparison (i.e., omnibus test) of the critical three-way interaction was non-significant ($\chi^2 = 4.56$, $p = 0.10$). However, the model estimates within the full model indicated a significant difference between the Exposure and Control groups ($\beta = 0.23$, $p = 0.04$); the difference between the Feedback and Control groups trended in the same direction but was non-significant ($\beta = 0.15$, $p = 0.16$). To better understand the three-way interaction, we created *post-hoc* models to directly compare performance by the Control and Exposure groups for the L2 accent condition at Pre-Test and then (in a separate model) at Post-Test. Model estimates indicated that at Pre-Test the Exposure group had (non-significantly) poorer performance ($\beta = -0.06$, $p = 0.19$) than the Control group, and (non-significantly) better performance ($\beta = 0.01$, $p = 0.77$) at Post-Test; critically, the size of these trends suggest that the three-way interaction was driven by a difference at Pre-Test, not Post-Test. Given that the omnibus test was non-significant, and the significant model estimate appears to have been driven by a Pre-Test difference, we conclude that no meaningful (training-related) differences emerged between the Control and training groups in the recognition accuracy dataset.

Response time data from visual categorization task

Response time data for all conditions is presented in Figure 2. We report all log-likelihood model comparisons in Table 3 and provide full model summaries in Supplemental materials. Matching the results of the recognition accuracy analysis, results of the response time analysis indicated improvement (i.e., reduction in response times) from Pre-Test to Post-Test ($\beta = -99.94$, $p < 0.001$). Improvements were also largest for the L2 accent condition (significant interaction of Condition and Session: $\chi^2 = 50.05$, $p < 0.001$). However, no differences in improvement emerged based on Group (all $ps > 0.05$).

Overall, participants had significantly slower response times on the secondary task when presented with an L2 accent in the primary task (as compared to an L1 accent; $\beta = 39.45$, $p < 0.001$). The response times did not differ overall by Group ($\chi^2 = 1.38$, $p = 0.50$), nor did the effect of Group interact with Condition ($\chi^2 = 0.01$, $p = 0.99$), Session ($\chi^2 = 3.03$, $p = 0.22$), or a combination of Condition and Session ($\chi^2 = 2.79$, $p = 0.25$). Model estimates of the three-way interactions were also non-significant, although the direction of the trends was as predicted: The difference in response times for the L1 and L2 accent conditions was reduced at Post-Test to a (non-significantly) larger degree for the Exposure ($\beta = -21.11$, $p = 0.16$) and Feedback ($\beta = -21.91$, $p = 0.14$) groups.

Training data

Transcription accuracy

Recognition accuracy data from the training sessions is presented in Figure 3. Fixed effects of the model included: Day (dummy-coded levels: Days 2, 3, 4), Group (dummy-coded levels: Control, Exposure, Feedback), as well as the interaction between Day and Group.
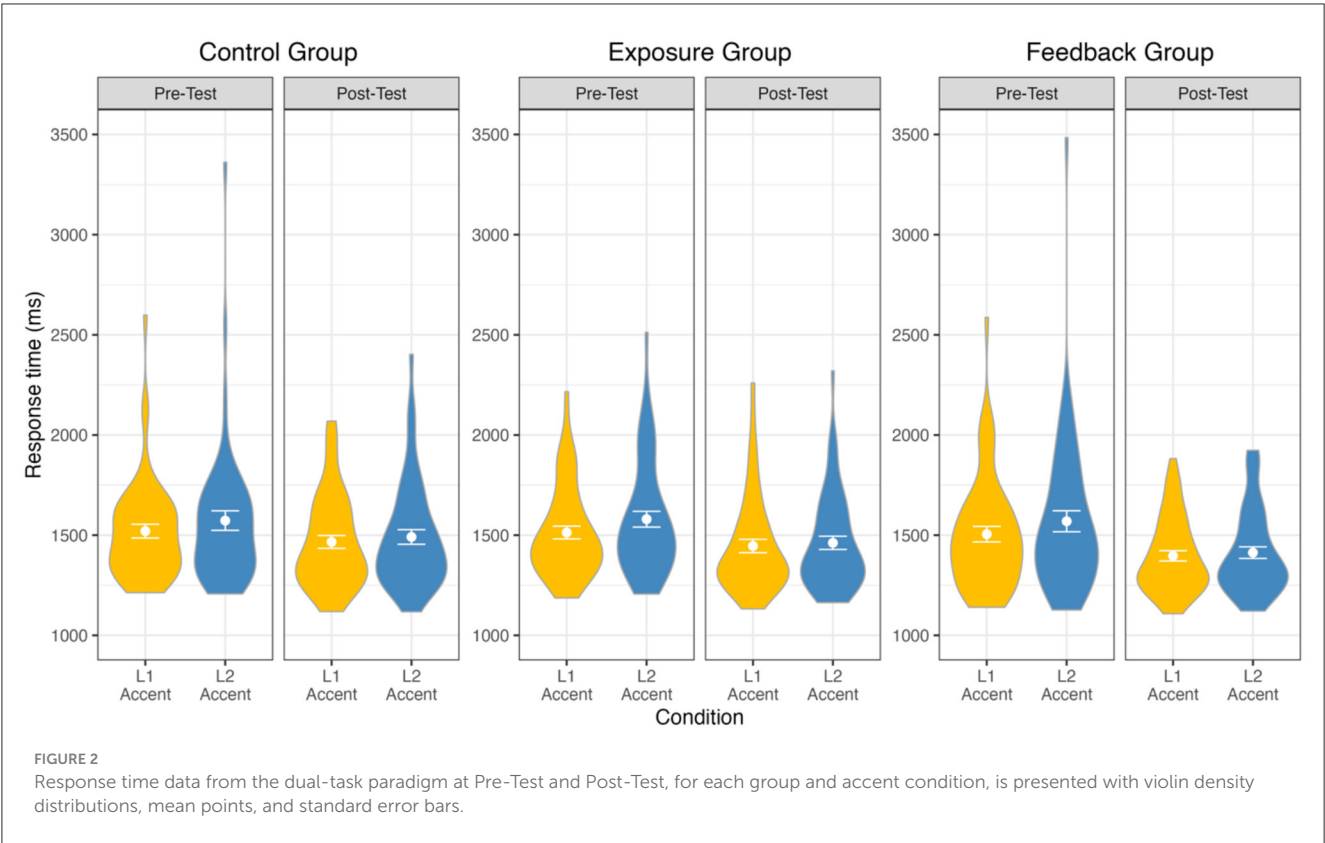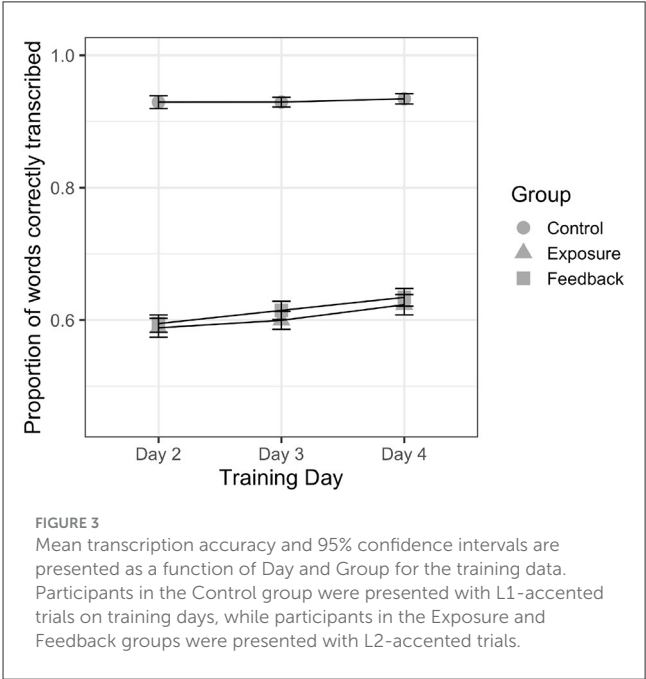
FIGURE 2
Response time data from the dual-task paradigm at Pre-Test and Post-Test, for each group and accent condition, is presented with violin density distributions, mean points, and standard error bars.

TABLE 3 Log-likelihood model comparisons from analyses of dual-task response time data.

Effect	χ^2	df	p
Condition	34.00	1	<0.001
Session	33.35	1	<0.001
Group	1.38	2	0.50
Condition: session	50.05	1	<0.001
Condition: group	0.01	2	0.99
Session: group	3.03	2	0.22
Condition: session: group	2.79	2	0.25

df, degrees of freedom.

FIGURE 3
Mean transcription accuracy and 95% confidence intervals are presented as a function of Day and Group for the training data. Participants in the Control group were presented with L1-accented trials on training days, while participants in the Exposure and Feedback groups were presented with L2-accented trials.

Log-likelihood model comparisons indicated that the effects of Day ($\chi^2 = 25.76$, $p < 0.001$) and Group ($\chi^2 = 496.50$, $p < 0.001$) both improved model fit. Model estimates revealed improvement across the training days, such that performance on Day 3 ($\beta = 0.05$, $p < 0.05$) and Day 4 ($\beta = 0.12$, $p < 0.001$) were each better than Day 2. Releveling of the fixed effect of Day in the model confirmed that the difference in performance between Days 3 and 4 was also significant ($\beta = 0.07$, $p = 0.002$). For the effect of Group, performance of subjects assigned to the Exposure and Feedback groups (both of which received entirely Mandarin-accented stimuli) was poorer than that of subjects assigned to the Control group (which received entirely American-accented stimuli; $ps < 0.001$). The fixed effect of Group was releveled in the model to directly compare the Exposure and Feedback groups but revealed no significant difference in overall performance ($\beta = 0.05$, $p = 0.22$). The interaction of Day and Group was non-significant ($\chi^2 = 3.20$, $p = 0.53$), indicating consistent improvement across days regardless of the assigned training.

Discussion

In the present study, we investigated whether brief daily exposure to unfamiliar L2 accent improves listeners' ability to accurately understand speech and, simultaneously, whether it reduces the cognitive load associated with speech processing. At Pre-Test and Post-Test, participants were presented with multiple L1- and L2-accented speakers while completing a dual-task paradigm. We predicted that response times (an index of cognitive load) during the L2 accent trials would be shortened (improved) for the subjects assigned to L2 accent training groups as compared to a Control group. Additionally, we predicted that speech recognition accuracy would improve in the L2 accent condition for the L2 accent training groups. Overall, our results indicated similar improvements for all groups. Critically, Post-Test performance for the L2 accent condition for the Control and L2 accent training groups did not differ significantly (although all trends were in the predicted directions). We conclude that the L2 accent trainings implemented in the present study did not successfully promote long-term learning benefits of a statistically meaningful magnitude. However, we also emphasize that the present effort is a methodologically informative starting point for future research on this topic.

Our examination of the data from the dual-task paradigm on Days 1 and 5 consistently revealed the following across all three random assignments: (1) Participants improved at the task from Pre- to Post-Test, making it easier both in terms of cognitive load (faster response times) and perceptual processing (better recognition accuracy); (2) Listening performance was poorer and cognitive load was greater for the L2 accent condition as compared the L1 accent condition. These outcomes were to be expected given the design of the experiment, and general participant learning effects. Of particular interest to the present study's aims was the interaction of these two elements with the training manipulation.

In the analysis of recognition accuracy, we found some evidence that the Exposure group (i.e., L2 accent training without implicit feedback), in particular, may have improved from Pre-Test to Post-Test to a larger degree than the Control group. However, *post-hoc* analyses following the critical three-way interaction revealed that the Exposure group likely demonstrated larger improvement because of a difference at Pre-Test, not Post-Test. In other words, it is impossible to determine whether they improved to a larger degree than the Control group because they had more "room to improve" or because of they received training with the L2 accent. Given that the omnibus test of the three-way interaction (i.e., including the Feedback group) was non-significant, and the corresponding model estimate for the Feedback group was non-significant, we conclude that the present study did *not* find sufficient evidence to indicate a benefit of the L2 accent trainings on listening performance. In this same vein, we also did not find evidence that indicated any benefit of the L2 accent training with implicit feedback over the L2 accent training without implicit feedback. Given prior evidence that the presentation of subtitles can promote adaptation to L2 accent (Chan et al., 2020), we had predicted that listeners in the Feedback group would show a larger benefit than listeners in the Exposure group. Our results suggest that presenting target sentences after listening (instead of in tandem with listening) may

not provide similar training benefits (cf., Burchill et al., 2018). In future research, a manipulation that presents lexical items in tandem with auditory targets may prove to have a larger effect.

For the response time data (i.e., cognitive load), analyses indicated similar trends: The difference in cognitive load for the L1 and L2 accent conditions was reduced on Day 5 compared to Day 1, but to a similar degree for all participant groups. One limitation of the present dataset may be the use of a reaction time task to measure cognitive load. Indeed, the amount of variance in subject response times may have reduced our power to detect the critical interaction. Measures of cognitive load (or "listening effort") can differ markedly in their sensitivity to detect differences; for example, examining the cognitive load associated with speech-in-noise perception, Strand et al. (2018) found that effect sizes were larger when using a semantic dual-task paradigm than a complex dual-task paradigm (the latter of which is most similar to the present study's paradigm). Pupillometry, a psychophysiological measure of cognitive load, was even more sensitive than these dual-task measures. In future work, using a psychophysiological measure, such as pupillometry or eye-tracking, may provide greater precision and ability to detect changes in cognitive load as well as processing speed.

As expected, participants presented with L1 accent (the Control group) on Days 2, 3, and 4 of the study had higher overall transcription accuracy on those days of the study than participants presented with L2 accent. Matching this outcome, participants in the L2 accent training groups also self-reported that the task was more effortful than participants in the Control group. Across days, all groups showed steady improvement in listening performance, matching prior work that has demonstrated sustained benefits of L2 accent trainings over brief periods (Lindemann et al., 2016; Xie et al., 2017). There was no difference, however, in the rate of improvement between the Control and L2 accent training groups. Additionally, we had predicted that the Feedback group may show more rapid improvement than the Exposure group, but this was not the case. Participants in the Exposure and Feedback groups did, however, perceive their performance as improving across days, whereas the Control group perceived their performance as declining. The Feedback group also perceived their performance as marginally more positive than the Exposure group, which may reflect their superior ability to self-assess performance with the implicit feedback available to them.

Self-reported motivation also varied by group. Participants in the Feedback group reported greater motivation to do well at the task than participants in the Control or Exposure groups. The Exposure and Control groups did not significantly differ, although the trends in the data suggested that both of the L2 accent training groups reported higher motivation than the Control group. It may be the case that the L2 accent stimuli were more engaging to listeners, albeit more challenging.

Limitations and future directions

We acknowledge the possibility that limited statistical power may have encumbered our ability to detect significant training benefits in the present study. In the response time data, in

particular, the degree of variance may have reduced our ability to detect effects. We suggest increasing the number of trials per condition in future work when comparing dual-task data across sessions or using a cross-modal matching task in place of a dual-task paradigm (also referred to as a "semantic" or "linguistic" dual-task paradigm; Strand et al., 2018). Although we created 100 novel stimuli in addition to the 200 SNST items, across 5 days this resulted in only 39 items per accent condition per task. In lieu of a larger set of sentence recordings, one solution would be to repeat items, particularly on training days, in future research (see Bradlow and Bent, 2008; Baese-Berk et al., 2013). In Pre- and Post-Test measures, it is critical to include novel stimuli in order to prevent item-specific learning effects from artificially inflating performance; however, analyses of training sessions are typically less crucial, and items could be repeated. There may even be a benefit to repeating items across training sessions, although to our knowledge this has yet to be examined directly. It is our hope that the present study can serve as a benchmark when selecting paradigms and estimating power in future investigations of multi-day accent trainings. We recommend that future studies maximize potential effect sizes via a combination of the following methods: (1) Using more sensitive measures of cognitive load, (2) Increasing the number of trials in test sessions, (3) Increasing the number of trials in training sessions, and (4) Increasing the number of participants.

With regard to the first recommendation, we predict based on prior evidence from speech-in-noise perception (Strand et al., 2018) that cross-modal matching tasks may produce larger effects than dual-task paradigms that involve a non-linguistic secondary task—although direct comparisons of the sensitivity of these paradigms for L2 accent perception/adaptation have yet to be conducted. Comparing the results of Clarke and Garrett (2004) with Brown et al. (2020), however, provides some indication of what types tasks may be most sensitive in the context of L2 accent adaptation: In Clarke and Garrett (2004), rapid (single session) adaptation to L2 accent was robustly demonstrated both across four experimental blocks (each containing 16 trials) and within the early trials. In that study, a cross-modal matching task was used; specifically, participants completed a task where they responded "yes" or "no" to visually-presented sentence-final probe words. In contrast, Brown et al. (2020) used the same non-linguistic dual-task paradigm as the present study and only found evidence of rapid adaptation within the first 20 trials, not across the full 50-trial session. Other differences between the two studies (type of L2 accent, presence of background noise, etc.) may account for the deviating outcomes, but we (cautiously) recommend based on outcomes of these prior studies that researchers may be best served with cross-modal matching tasks in future work. We can also (more confidently) recommend pupillometry as a measure of cognitive load, which proved to be more sensitive than the dual-task paradigms in both Strand et al. (2018) and Brown et al. (2020).

From Pre- to Post-Test, (non-significant) trends in the data indicated larger benefits for the measure of listening performance than the measure of cognitive effort. This outcome runs counter to the findings of Bieber and Gordon-Salant (2021), in which benefits were observed at a test session 1-week after training for cognitive load but not listening performance. One possible explanation for the contrary outcome in the present study may

be the design of the middle (training) days, which utilized a speech transcription task rather than the dual-task paradigm from the Pre- and Post-Test sessions. Thus, participants received more extensive training with the linguistic task, but not the non-linguistic (visual) task, from the dual-task paradigm. It may be the case that these training sessions were better situated to promote near-transfer to the more similar (linguistic) task. In future work, matching the designs of training and test sessions may be ideal to remove any potential differential transfer effects by task type.

One strength of the present study was the inclusion of L2-accented stimuli presented in quiet, as opposed to in noise. Although adding noise to stimuli can make it easier to match intelligibilities across conditions (e.g., matching L1 to L2 speakers), prior evidence also indicates that the cognitive and/or perceptual resources recruited to support noisy vs. accented listening conditions may differ (McLaughlin et al., 2018). Thus, when examining questions pertaining to the perception of L2 accent, using L2-accented stimuli presented in noise may not always be suitable. To prevent ceiling effects in the present study, we decided to use semantically-anomalous sentences (e.g., "the wrong shot led the farm"), which pose a different potential issue: Namely, anomalous sentences reduce a listener's ability to use top-down information during speech processing, and are therefore less ecologically valid. Studies that use these types of items thus give a more direct assessment of bottom-up processing at the cost of limited generalizability of the findings. In future work, focusing solely on measures of cognitive load (as opposed to a combination of cognitive load and intelligibility measures) can remove these types of obstacles and allow for more ecological examinations of accent accommodation.

Conclusion

Although L2 accent can pose a challenge during speech processing, listeners are able to rapidly accommodate L2 speakers' unique productions, thereby reducing cognitive load (Clarke and Garrett, 2004; Brown et al., 2020). Additionally, correlational evidence suggests that the efficiency and accuracy of L2 accent processing depends on a listener's prior (real world) experience, with more experienced listeners typically processing L2 accent faster and more accurately (Kennedy and Trofimovich, 2008; Porretta et al., 2020). Empirical evidence connecting these two literatures, however, is lacking: Few studies to date have examined perceptual accommodation of L2 accent across multiple days (or weeks, etc.). In the present study, we took a first step toward filling this empirical gap, implementing a dual-task paradigm to measure changes in cognitive load and listening performance for perception of L2 accent across a 5-day period. Participants were either exposed to the L1 (Control) or L2 accent in the interim days, and half of the subjects exposed to L2 accent were provided with implicit feedback. Our results did not show a benefit of the L2 accent trainings, despite a larger sample size ($n > 50$ per group) than prior work (although all trends were in the predicted directions). We conclude that the L2 accent

trainings implemented in the present study did not successfully promote long-term learning benefits of a statistically meaningful magnitude, but also emphasize that the present effort is a methodologically informative starting point for future research on this topic.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://osf.io/2qz5n/files/osfstorage.

Ethics statement

The studies involving humans were approved by Washington University in St. Louis IRB. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

DM, MB-B, and KV contributed to conception and design of the study. DM performed the statistical analysis and wrote the first draft of the manuscript. All authors contributed to manuscript revision, read, and approved the submitted version.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/flang.2024.1243678/full#supplementary-material

References

Arbuthnott, K., and Frank, J. (2000). Trail making test, part B as a measure of executive control: validation using a set-switching paradigm. *J. Clin. Exp. Neuropsychol.* 22, 518–528. doi: 10.1076/1380-3395(200008)22:4;1-0;FT518

Baese-Berk, M. M., Bradlow, A. R., and Wright, B. A. (2013). Accent-independent adaptation to foreign accented speech. *J. Acoust. Soc. Am.* 133, EL174–EL180. doi: 10.1121/1.4789864

Bates, D. M. M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* 67, 1–48. doi: 10.18637/jss.v067.i01

Bent, T., and Baese-Berk, M. (2021). "Perceptual learning of accented speech," in *The Handbook of Speech Perception*, eds. J. S. Pardo, L. C. Nygaard, R. E. Remez and D. B. Pisoni. doi: 10.1002/9781119184096.ch16

Bieber, R. E., and Gordon-Salant, S. (2021). Improving older adults' understanding of challenging speech: auditory training, rapid adaptation and perceptual learning. *Hear. Res.* 402:108054. doi: 10.1016/j.heares.2020.108054

Bradlow, A. R., and Bent, T. (2008). Perceptual adaptation to non-native speech. *Cognition* 106, 707–729. doi: 10.1016/j.cognition.2007.04.005

Brown, V. A., McLaughlin, D. J., Strand, J. F., and Van Engen, K. J. (2020). Rapid adaptation to fully intelligible nonnative-accented speech reduces listening effort. *Q. J. Exp. Psychol.* 73, 1431–1443. doi: 10.1177/1747021820916726

Burchill, Z., Liu, L., and Jaeger, T. F. (2018). Maintaining information about speech input during accent adaptation. *PLoS ONE* 13:e0199358. doi: 10.1371/journal.pone.0199358

Chan, K. Y., Lyons, C., Kon, L. L., Stine, K., Manley, M., and Crossley, A. (2020). Effect of on-screen text on multimedia learning with native and foreign-accented narration. *Learn. Instruct.* 67:101305. doi: 10.1016/j.learninstruc.2020.101305

Clarke, C. M., and Garrett, M. F. (2004). Rapid adaptation to foreign-accented English. *J. Acoust. Soc. Am.* 116, 3647–3658. doi: 10.1121/1.1815131

Derwing, T. M., Rossiter, M. J., and Munro, M. J. (2002). Teaching native speakers to listen to foreign-accented speech. *J. Multiling. Multicult. Dev.* 23, 245–259. doi: 10.1080/01434630208666468

Dietrich, S., and Hernandez, E. (2022). *Language Use in the United States: 2019.* Suitland, MD: American Community Survey Reports.

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychol. Rev.* 105, 251–279. doi: 10.1037/0033-295x.105.2.251

Johnson, K. (1997). "Speech perception without speaker normalization: an exemplar model," in *Talker Variability in Speech Processing*, eds. K. Johnson and J. Mullenix (Academic Press), 145–166.

Kennedy, S., and Trofimovich, P. (2008). Intelligibility, comprehensibility, and accentedness of L2 speech: the role of listener experience and semantic context. *Can. Modern Lang. Rev.* 64, 459–489. doi: 10.3138/cmlr.64.3.459

Kleinschmidt, D. F., and Jaeger, T. F. (2015). Robust speech perception: recognize the familiar, generalize to the similar, and adapt to the novel. *Psychol. Rev.* 122, 148–203. doi: 10.1037/a0038695

Kuznetsova, A., Brockhoff, P. B., and Christensen, R. H. B. (2017). lmerTest package: tests in linear mixed effects models. *J. Stat. Softw.* 82, 1–26. doi: 10.18637/jss.v082.i13

Lindemann, S., Campbell, M. A., Litzenberg, J., and Close Subtirelu, N. (2016). Explicit and implicit training methods for improving native English speakers? comprehension of nonnative speech. *J. Second Lang. Pronunc.* 2, 93–108. doi: 10.1075/jslp.2.1.04lin

MacLeod, C. M. (1991). Half a century of research on the Stroop effect: an integrative review. *Psychol. Bull.* 109:163. doi: 10.1037/0033-2909.109.2.163

Magnuson, J. S., Nusbaum, H. C., Akahane-Yamada, R., and Saltzman, D. (2021). Talker familiarity and the accommodation of talker variability. *Attent. Percept. Psychophys.* 83, 1842–1860. doi: 10.3758/s13414-020-02203-y

McLaughlin, D. J., Baese-Berk, M. M., Bent, T., Borrie, S. A., and Van Engen, K. J. (2018). Coping with adversity: individual differences in the perception of noisy and accented speech. *Attent. Percept. Psychophys.* 80, 1559–1570. doi: 10.3758/s13414-018-1537-4

Newman, R. S., and Evers, S. (2007). The effect of talker familiarity on stream segregation. *J. Phonet.* 35, 85–103. doi: 10.1016/j.wocn.2005.10.004

Nye, P. W., and Gaitenby, J. H. (1974). The intelligibility of synthetic monosyllabic words in short, syntactically normal sentences. *Haskins Lab. Status Rep. Speech Res.* 38:43.

Nygaard, L. C., and Pisoni, D. B. (1998). Talker-specific learning in speech perception. *Percept. Psychophys.* 60, 355–376. doi: 10.3758/BF03206860

Pierrehumbert, J. (2001). "Lenition and contrast," in *Frequency and the Emergence of Linguistic Structure*, 137.

Porretta, V., Buchanan, L., and Järvikivi, J. (2020). When processing costs impact predictive processing: the case of foreign-accented speech and accent experience. *Attent. Percept. Psychophys.* 82, 1558–1565. doi: 10.3758/s13414-019-01946-7

Porretta, V., and Tucker, B. V. (2019). Eyes wide open: pupillary response to a foreign accent varying in intelligibility. *Front. Commun.* 4:8. doi: 10.3389/fcomm.2019.00008

R Core Team (2021). *R: A Language and Environment for Statistical Computing.* Vienna: R Foundation for Statistical Computing. Available online at: https://www.R-project.org/ (accessed January 1, 2022).

Sidaras, S. K., Alexander, J. E., and Nygaard, L. C. (2009). Perceptual learning of systematic variation in Spanish-accented speech. *J. Acoust. Soc. Am.* 125, 3306–3316. doi: 10.1121/1.3101452

Smith, S. L., Pichora-Fuller, M. K., and Alexander, G. (2016). Development of the word auditory recognition and recall measure: a working memory test for use in rehabilitative audiology. *Ear Hear.* 37, e360–e376. doi: 10.1097/AUD.0000000000000329

Strand, J. F., Brown, V. A., Merchant, M. B., Brown, H. E., and Smith, J. (2018). Measuring listening effort: convergent validity, sensitivity, and links with cognitive and personality measures. *J. Speech Lang. Hear. Res.* 61, 1463–1486. doi: 10.1044/2018_JSLHR-H-17-0257

Witteman, M. J., Weber, A., and McQueen, J. M. (2013). Foreign accent strength and listener familiarity with an accent codetermine speed of perceptual adaptation. *Attent. Percept. Psychophys.* 75, 537–556. doi: 10.3758/s13414-012-0404-y

Xie, X., Jaeger, T. F., and Kurumada, C. (2023). What we do (not) know about the mechanisms underlying adaptive speech perception: a computational framework and review. *Cortex* 66, 377–424. doi: 10.1016/j.cortex.2023.05.003

Xie, X., Theodore, R. M., and Myers, E. B. (2017). More than a boundary shift: Perceptual adaptation to foreign-accented speech reshapes the internal structure of phonetic categories. *J. Exp. Psychol.* 43:206. doi: 10.1037/xhp0000285