Original Article

QJEP

Quarterly Journal of Experimental Psychology
1–13

DOI: 10.1177/1747021820916726
qjep.sagepub.com
SAGE

Rapid adaptation to fully intelligible nonnative-accented speech reduces listening effort

Violet A Brown[1], Drew J McLaughlin[1], Julia F Strand[2] and Kristin J Van Engen[1]

Abstract

In noisy settings or when listening to an unfamiliar talker or accent, it can be difficult to understand spoken language. This difficulty typically results in reductions in speech intelligibility, but may also increase the effort necessary to process the speech even when intelligibility is unaffected. In this study, we used a dual-task paradigm and pupillometry to assess the cognitive costs associated with processing fully intelligible accented speech, predicting that rapid perceptual adaptation to an accent would result in decreased listening effort over time. The behavioural and physiological paradigms provided converging evidence that listeners expend greater effort when processing nonnative- relative to native-accented speech, and both experiments also revealed an overall reduction in listening effort over the course of the experiment. Only the pupillometry experiment, however, revealed greater adaptation to nonnative- relative to native-accented speech. An exploratory analysis of the dual-task data that attempted to minimise practice effects revealed weak evidence for greater adaptation to the nonnative accent. These results suggest that even when speech is fully intelligible, resolving deviations between the acoustic input and stored lexical representations incurs a processing cost, and adaptation may attenuate this cost.

Keywords

Speech perception; listening effort; accented speech; dual-task; pupillometry

Received: 28 August 2019; revised: 13 February 2020; accepted: 14 February 2020

Understanding speech is a perceptually and cognitively demanding task that requires listeners to map highly variable acoustic input onto linguistic representations in their mental lexicons. This variability in the acoustic input can certainly hinder intelligibility (Bradlow & Bent, 2008; Nygaard et al., 1994), but also has more general cognitive consequences. A growing body of research suggests that despite the subjective ease with which spoken word recognition occurs, it does not occur effortlessly, and instead requires the recruitment of cognitive resources, often referred to as *listening effort*—the cognitive resources a listener expends to complete a listening task (see Downs, 1982; Pichora-Fuller et al., 2016). The concept of listening effort relies on the assumption that humans have a finite cognitive capacity for processing information, and increasing task difficulty or simultaneously engaging in multiple tasks puts stress on this limited-capacity system (Kahneman, 1973). Thus, listening to and understanding a spoken utterance requires greater effort when the listening conditions are more difficult—for example, when the level of the background noise is increased (Downs & Crum, 1978; Picou & Ricketts, 2014; Strand et al., 2018), when the speech is produced in a conversational rather than listener-oriented style (Van Engen et al., 2012), or when listening to nonnative- relative to native-accented speech (Van Engen & Peelle, 2014; see Mattys et al., 2012, for a review of speech recognition in adverse conditions).

A commonly used method of quantifying listening effort that capitalises on the concept of a limited-capacity system is the dual-task paradigm, which requires participants to

---

[1]Department of Psychological & Brain Sciences, Washington University in St. Louis, Saint Louis, MO, USA
[2]Department of Psychology, Carleton College, Northfield, MN, USA

Corresponding author:
Violet A Brown, Department of Psychological & Brain Sciences, Washington University in St. Louis, 1 Brookings Drive, Saint Louis, MO 63130-4899, USA.
Email: violet.brown@wustl.edu

recognise speech while performing a concurrent cognitive task (e.g., Alsius et al., 2007; Gagné et al., 2017; Picou & Ricketts, 2014). In dual-task paradigms, participants are typically told to prioritise the speech task while simultaneously completing the secondary task to the best of their abilities. Thus, as the difficulty of the speech recognition task increases, fewer resources remain to quickly and accurately perform the secondary task, and increases in listening effort are assumed to be reflected by slower response times to the secondary task (Bourland-Hicks & Tharpe, 2002 ; Desjardins & Doherty, 2013; Downs, 1982; Fraser et al., 2010; Strand et al., 2018).

In addition to behavioural measures of listening effort, researchers often use physiological measures to quantify the extent to which engagement in a speech task requires cognitive resources. Here, cognitive load (defined as the degree to which a task's demands consume the cognitive resources available at a given moment; Pichora-Fuller et al., 2016) is thought to be associated with increased stress and arousal, which is reflected in changes in heart rate variability (Seeman & Sims, 2015), electromyographic activity (Mackersie & Cones, 2011), and skin conductance (Mackersie & Cones, 2011; Seeman & Sims, 2015). Recruitment of prefrontal and premotor cortices, as well as the cingulo-opercular network, may also be indicative of increased listening effort (see Peelle, 2018).

The most commonly used physiological measure of listening effort is *pupillometry*, which measures the change in pupil size over time (for a review, see Van Engen & McLaughlin, 2018). This method has also been used as an index of cognitive load more generally (Beatty, 1982; Klingner et al., 2011). A substantial body of work has demonstrated that pupil size systematically increases as the intelligibility of speech decreases (Kuchinsky et al., 2013; McGarrigle et al., 2016; Porretta & Tucker, 2019; Zekveld et al., 2010), and as degradation of the speech signal increases, even when the degradation does not affect intelligibility (Winn et al., 2015). The pupillary response to changes in task demands is often referred to as the *task-evoked pupillary response* (Beatty, 1982; referred to simply as *pupil response* from here forward). Given the temporal sensitivity of the pupil response, pupillometry is particularly useful for studying speech processing, as it allows researchers to track rapid changes in listening effort that may unfold over the course of an utterance.

Listening effort is most often induced in laboratory settings by increasing the level of the background noise, but other conditions or listener characteristics that interfere with speech intelligibility may also increase listening effort, such as noise vocoding (Winn et al., 2015), hearing impairment (McCoy et al., 2005), or listening to speech in a second language (Borghini & Hazan, 2018). Compared with processing speech produced by a native-accented talker, processing nonnative-accented speech also appears to increase listening effort. In a recent pupillometry study, McLaughlin and Van Engen (2020) found that native English-speaking participants show larger increases in pupil dilation—indicating greater listening effort—when listening to Mandarin Chinese–accented English relative to standard American English sentences. Notably, the stimuli used in that study were fully intelligible (i.e., all four keywords in the sentences included in the primary analyses were correctly identified by participants), suggesting that even when speech intelligibility is at ceiling, resolving mismatches between the accented speech input and representations of words in the listener's mental lexicon may require additional cognitive resources (e.g., working memory, attention; Van Engen & Peelle, 2014). Further, using monosyllabic words recorded by multiple Mandarin Chinese–accented speakers of English, Porretta and Tucker (2019) found that pupil dilation is larger for less intelligible accented speech. Thus, it appears that nonnative-accented speech requires additional listening effort to process, and this processing cost increases as intelligibility decreases.

Although nonnative-accented speech may initially be processed more slowly and less accurately than native speech, listeners adapt to nonnative-accented speech quite rapidly (Clarke & Garrett, 2004). Further, experience with a particular nonnative accent can facilitate recognition of words produced by a different talker with the same accent (Alexander & Nygaard, 2019; Bradlow & Bent, 2008; Sidaras et al., 2009), and adaptation effects appear to remain after short delay periods (Kraljic & Samuel, 2005). One explanation for these accent adaptation effects involves modifications to the ways in which the acoustic signal is mapped onto phonemic categories stored in memory. It may be that prior to adaptation to an accent, the incoming speech sounds do not match stored representations of phonemes, and the resulting incorrect phoneme identification may cause the word recognition system to essentially backtrack to correct its mistakes. Thus, perceptual adaptation may improve intelligibility by adjusting how speech sounds are mapped onto representations in memory (see Van Engen & Peelle, 2014). If this is the case, then these perceptual adjustments may reduce listening effort by reducing the number of times incorrect mappings from sounds to phonemes must be inhibited and reevaluated. This interpretation is in line with the general claims of the Ease of Language Understanding (ELU) model, which states that when the speech signal does not match representations in memory, additional cognitive resources are employed to resolve the mismatch (see Rönnberg et al., 2019 for an updated version of the 2008 and 2010 models; Rudner et al., 2011).

To summarise, perceptual adaptation to a nonnative accent improves identification of words and sentences produced in that accent. It is not yet known, however, whether adaptation to nonnative-accented speech occurs for listening effort when intelligibility is at ceiling. One possibility is that even after near-perfect accuracy is attained for understanding nonnative-accented speech, differences in listening effort persist (McLaughlin & Van Engen, 2020), and effort does not reduce with exposure. A growing body of research indicates that differences in listening effort may be observed even when intelligibility is equated. For example, though widely used, noise-reduction algorithms in hearing aids typically do not improve speech recognition in background noise, yet they reduce the effort necessary to recognise speech (Desjardins & Doherty, 2014; Sarampalis et al., 2009). Dissociations between effort and intelligibility have also been demonstrated by changing the demands of the speech task (Mackersie & Cones, 2011). Thus, it may be that adaptation occurs for intelligibility but not for listening effort.

Another possibility is that additional cognitive resources are initially required to recognise and encode speech produced in an unfamiliar nonnative accent (see Van Engen & Peelle, 2014, for a review), but after sufficient exposure these additional resources become less necessary because listeners have adjusted their phoneme categories—or adjusted the ways in which the acoustic input is mapped onto phonemic representations—such that the words produced by the nonnative-accented talker are more efficiently mapped onto representations of words in their mental lexicons. This view is supported by several studies showing that accented speech (both nonnative and regional) is processed more slowly than native speech, but this cost can be rapidly attenuated such that response times to nonnative-accented speech are nearly identical to those to native-accented speech after adaptation (Clarke & Garrett, 2004). An extension of this possibility is that exposure to a nonnative accent reduces listening effort, but after adaptation a difference in listening effort between nonnative- and native-accented speech still persists.

To date, no research has examined how perceptual adaptation to an accent affects listening effort. This study therefore employed a multimethod approach to investigate this relationship. Experiment 1 used the complex dual-task, a commonly used dual-task paradigm in the listening effort literature (Brown & Strand, 2018; Picou & Ricketts, 2014; Sarampalis et al., 2009; Strand et al., 2018). Experiment 2 used a subset of the speech materials in Experiment 1, and listening effort was quantified using pupillometry. We opted to include both measures of listening effort for two reasons. First, several recent papers have raised concerns that different measures of listening effort may not in fact be measuring the same underlying construct (Alhanbali et al., 2019; Strand et al., 2018), so including multiple paradigms will help us evaluate whether any observed effects are specific to the measure of effort employed. Further, paradigms differ in their ability to detect changes in listening effort resulting from changes in the level of the background noise (e.g., Johnson et al., 2015; Picou & Ricketts, 2014), so including both a dual-task paradigm and pupillometry enabled us to assess whether these measures also differ in their sensitivity to changes in listening effort associated with native versus nonnative speech, as well as adaptation to nonnative-accented speech.

We hypothesised that even when intelligibility is matched, nonnative-accented speech would require greater listening effort relative to native-accented speech as measured by both a dual-task paradigm (as indicated by slower response times to the secondary task; Experiment 1) and pupillometry (as indicated by greater pupil dilation; Experiment 2). Further, we hypothesised that adaptation to nonnative-accented speech would reduce the listening effort required to process that accent to a greater extent than for native-accented speech. This would indicate that adaptation to nonnative-accented speech occurs above and beyond adaptation to the task or the particular talker's voice.

Experiment 1

All hypotheses, analyses, and exclusion criteria, as well as the sample size justification for Experiment 1, were pre-registered and are available at https://osf.io/3vhjw. Deviations from our pre-registered plan are explicitly noted in the article. Data, materials, and code for the analyses reported in this experiment as well as in Experiment 2 are available at https://osf.io/xajdw/.

Method

Participants. College-aged native English speakers (aged 18–23) with self-reported normal hearing and normal or corrected-to-normal vision were recruited from the Washington University in St. Louis psychology subject pool. We pre-registered a sample size of 80, and we recruited 103 participants to attain this sample size after excluding participants who met the pre-registered exclusion criteria. Participants completed a language background questionnaire before participating in the experiment, and before analysing any data, participants were excluded if they reported speaking with a Mandarin accent, having substantial previous exposure to Mandarin-accented English, having parents or close family members who speak with a Mandarin accent, or having taken Mandarin in a classroom setting. These questionnaire-related criteria eliminated 14 participants, as well as two additional participants who filled out the form incorrectly (which made it difficult for the authors to discern whether they had substantial exposure to Mandarin-accented English). Four additional participants were excluded for having poor accuracy on the sentence recognition task, one was excluded for having poor accuracy at the secondary response time task, and one was excluded for having average response times more than three standard deviations above the mean for at least one condition. This resulted in usable data from 81 participants, so before performing any other analyses, we removed the last participant that had been run in the counterbalanced order with 41 rather than 40 participants (to enable a fully balanced design). The Washington University in St. Louis Institutional Review Board approved all research procedures. Participants in this experiment and in Experiment 2 received course credit for participation, with the exception of eight participants in this experiment who received $5 for 30 min of participation because the research subject pool had closed for the summer. This experiment took approximately 30 min to complete.

Speech stimuli. Speech stimuli consisted of 160 sentences produced by two talkers. One talker ("native") was a standard American English speaker, and the other talker ("nonnative") was a Mandarin Chinese–accented speaker of English. The sentences contained four keywords each (e.g., "the *hot sun warmed* the *ground*") and were designed to contain high-frequency words that could be easily identified by nonnative speakers (Van Engen et al., 2012).

The original recordings of native and nonnative speech were not matched for length (mean for the nonnative talker: 2,163 ms; mean length for the native talker: 1,673 ms). To avoid confounding speech rate with accent, we matched the native and nonnative sentences for length by first reducing long silences between words in the nonnative speech files and then adjusting the length of both the native and nonnative speech files using the "stretch and pitch" process in Adobe Audition (version 10). We increased the speed of the nonnative talker's speech (decreased the duration of the audio files) by making the nonnative speech 88.67% of its original length, and decreased the speed of the native talker's speech (increased the duration of the audio files) by making the native speech 114.64% of its original length. Native and nonnative sentences were then matched on total root-mean-square amplitude using Adobe Audition. Speech stimuli were presented to participants binaurally through Beyer dynamic DT100 headphones with an Aphex HeadPod Model 454 high output headphone amplifier at a comfortable listening level without background noise.

Procedure. Participants were seated in a testing room in front of a 21.5-in. iMac computer. Stimulus presentation was controlled via SuperLab (Cedrus, version 5). Participants were asked to complete a primary sentence recognition task while also performing a secondary visual response time task, and were told to prioritise the speech task but attempt to perform both tasks to the best of their abilities. The dual-task paradigm we employed was the complex dual-task, which involves simultaneously listening to speech while responding to visually presented numbers on the screen (Brown & Strand, 2018; Picou & Ricketts, 2014; Sarampalis et al., 2009; Strand et al., 2018). On each trial, the onset of a sentence coincided with the presentation of two empty square boxes (measuring approximately 5 cm), one on the left and one on the right. After a variable delay of 600–800 ms (in 50 ms increments), a number between 1 and 8 appeared in one of the two boxes. This interval between the onset of the sentence for the primary speech task and the appearance of the target stimulus for the secondary task was chosen to ensure that participants would begin listening to and attempting to identify the sentence before the secondary task began. Given the duration of the sentences and the time it takes for participants to respond to the numbers, this means that most of the processing associated with the complex dual-task and sentence task occurred simultaneously.

Participants responded to the numbers on the screen by pressing one of two buttons on a button box (Cedrus RB-600) as quickly and accurately as possible, one with a left-facing arrow and one with a right-facing arrow. If the number on the screen was even (i.e., 2, 4, 6, 8), participants were instructed to respond by pressing the key with the arrow that points towards the side of the screen on which the number appeared (e.g., if a "2" appeared in the left box, participants should press the left button). If the number was odd, participants responded by pressing the key with the arrow that points away from the box with the number (e.g., if a "3" appeared in the left box, participants should press the right button).

In addition to responding to the visually presented numbers, participants were asked to repeat aloud the sentence they heard, and were told to guess when unsure. After the participant responded, there was an inter-sentence interval randomly selected from 3,000, 3,500, or 4,000 ms prior to presentation of the next sentence. The native- and nonnative-accented sentences were blocked and the order of the accent blocks was counterbalanced across two groups of participants. The same 160 sentences were used in the native and nonnative conditions, but for half of the participants a given sentence appeared in the native condition, and for the other half that sentence appeared in the nonnative condition. Thus, each sentence appeared in both conditions, but each participant heard each sentence only once. Within each block, sentences were presented in a randomised order. Prior to beginning the main task, participants completed 15 s of practice on the secondary number task, followed by six practice sentences with both the speech and number tasks. The speech stimuli used in the practice trials were recorded by a different standard American English speaker ("native") and a Korean-accented English speaker ("nonnative"). The accents were intermixed during the practice trials. We intentionally chose stimuli from a different nonnative accent during the nonnative practice trials to ensure that participants had not already begun adapting to the Mandarin accent by the time they began the experiment.

Results and discussion

The analyses for Experiment 1 employed linear mixed effects models via the *lme4* package (version 1.1.21; Bates et al., 2014) in R (version 3.5.2; R Core Team, 2016), and model comparisons were conducted via likelihood ratio tests to determine the significance of effects of interest. Where appropriate, p values for model parameters were obtained with the *lmerTest* package (version 3.1.0; Kuznetsova et al., 2017) using the Satterthwaite method for estimating the degrees of freedom. Following the recommendations of Barr and colleagues (2013), we attempted to utilise the maximal random effects structure justified by the design. When necessary, we included control parameters to help enable convergence, and removed random effects that contributed little to the total variance and/or had correlations with other random effects of 1.00 or −1.00 (which may indicate overfitting). In all analyses, participants and items (sentences) were included as random effects. Accent type was dummy-coded such that the native accent was the reference level, and trial number was treated as a continuous predictor. The dependent variable was response time to the secondary visual number task.

Prior to analysing the data, 156 trials with extreme response times were removed for meeting our pre-registered exclusion criterion that response times must fall within three median absolute deviations (MADs) of a participant's median response time for a given accent (native or nonnative). We had originally planned to remove trials with response times slower than 2,000 ms, but opted to remove that criterion before analysing the data because a fixed cut-off that is applied to both accent conditions would result in greater data loss in the condition with slower response times. We only analysed response time data from trials during which participants identified all four keywords correctly (96.53% of the data). This decision was made to ensure that any differences in effort that emerge between native and nonnative speech are not attributable to differences in intelligibility. Accuracy at the speech recognition task was 93.84% in the nonnative condition and 99.20% in the native condition. In the analyses reported below, we only analysed trials during which participants correctly classified the number in the secondary task as odd or even (96.91% of trials). A total of 11,839 trials were included in the analyses reported below: 5,748 trials in the nonnative condition and 6,091 trials in the native condition.

We first built a model that included accent and trial number as fixed effects and compared it to reduced models lacking either accent or trial number. Likelihood ratio tests indicated significant main effects of accent ($\chi_1^2 = 13.99$; $p < .001$) and trial number ($\chi_1^2 = 21.91$; $p < .001$). Examination of the summary output for the model containing both fixed effects indicated that response times to the secondary number task were on average an estimated 70 ms slower in the nonnative compared to the native condition ($\beta = 70.28$, $SE = 18.09$, $t = 3.89$, $p < .001$). This finding suggests that native English speakers expended more listening effort when recognising fully intelligible speech produced by a nonnative Mandarin-accented English speaker relative to a native English speaker. Furthermore, for every trial, response times to the unrelated secondary task decreased by an estimated 1 ms ($\beta = -1.11$, $SE = 0.22$, $t = -4.99$, $p < .001$), suggesting that listeners expended less listening effort as they adapted to the talker's voices. To quantify goodness-of-fit, we calculated the marginal and conditional R^2 values via the *MuMIn* package (version 1.43.6; Bartoń, 2019), which are based on equations and code from Nakagawa and Schielzeth (2013) and revised statistics based on Nakagawa et al. (2017). The linear mixed model (LMM) including both accent and trial number had a marginal R^2 value of $R^2_{LMM(m)} = .01$ (indicative of a very small effect of accent and trial number) and a conditional R^2 value of $R^2_{LMM(c)} = .53$. The value for $R^2_{LMM(m)}$ represents the variance explained only by the fixed effects, and $R^2_{LMM(c)}$ represents the variance explained by the fixed and random effects. Thus, the fixed effects alone explain very little variance in response times, but when participant- and item-level variability are taken into account, the model explains approximately 53% of the variance in response times.

Next, we assessed whether participants adapted to native and nonnative speech at different rates by testing the interaction between accent and trial number. This would indicate that adaptation to nonnative-accented speech occurs above and beyond adaptation to the specific talker's voice. We built a model with accent, trial number, and an accent-by-trial number interaction as fixed effects, and compared it to a reduced model lacking the interaction term. A likelihood ratio test indicated that this interaction was not significant ($\chi_1^2 = 1.51$; $p = .22$), suggesting that the adaptation effect was consistent across the native and nonnative conditions (see Figure 1).

Although the main effect of trial number indicated that response times decreased over the course of the experiment—a finding that would be expected if adaptation to the talker occurred—it is unclear whether this effect reflects adaptation or practice effects. The secondary visual response time task is certainly difficult at first, but it may be that participants quickly reach a high level of performance at that task, which could mask any subtle adaptation effects that may be occurring. That is, it may be that there is an interaction between accent and trial number, but the effects happen (and dissipate) so quickly that they are not detectable when examining the full dataset. Visual inspection of Figure 1 suggests that response times in the nonnative condition began to plateau around the 20th trial, so we conducted an exploratory analysis in which we subsetted the data and tested for the interaction in only the first 20 trials. In an attempt to minimise the influence of any practice effects, we performed this exploratory analysis on the first 20 trials within only the second block that each participant completed (either native or nonnative). This analysis allowed us to assess whether the hypothesised interaction between accent and trial number existed when practice effects were

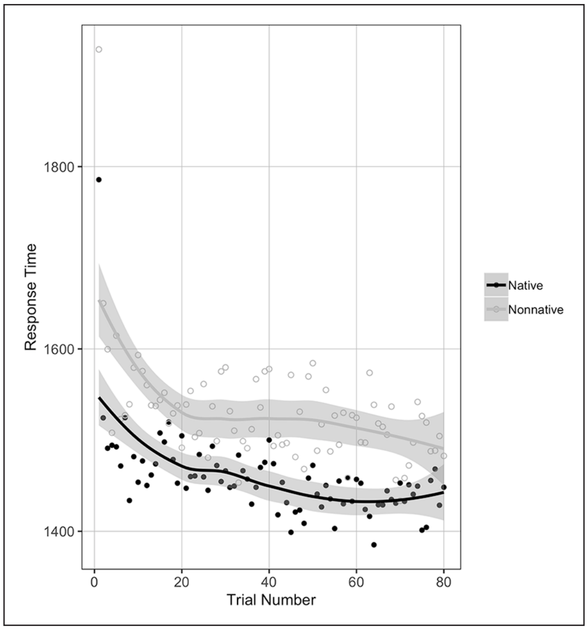

Figure 1. Scatterplot showing mean response times to the secondary visual response time task for each participant in each accent condition. The lines represent loess curves—smoothed local regression lines of best fit. Note that these curves were used for visualisation purposes only. Given that we were interested in the time course of adaptation to inform the trials we selected for use in the exploratory analysis described below, we opted to use a loess curve rather than the line produced by our model to depict the trend that adaptation happens most rapidly in the first ~20 trials more clearly.

reduced, because during the second block participants have already completed 80 trials and practice effects should therefore have already subsided. Note that because only the second block for each participant was used in this analysis, accent block was a between-subjects variable. We therefore did not include by-participant random slopes for accent in the model specification for the exploratory analysis.

A model using this subset of the data that included accent, trial number, and the critical accent-by-trial number interaction provided a better fit for the data than a model without the interaction ($\chi^2_1 = 4.12$; $p = .042$). Examination of the summary output for the full model indicated that response times decreased more rapidly across the first 20 trials of the second block for the nonnative relative to the native accent ($\beta = -7.32$, $SE = 3.61$, $t = -2.03$, $p = .046$). The marginal R^2 value for this exploratory model including the interaction term was $R^2_{LMM(m)} = .05$ and the conditional R^2 value was $R^2_{LMM(c)} = .59$. Thus, we found weak evidence that participants adapt more quickly to nonnative than native speech in the early stages of exposure (Figure 2). However, it is important to note that the number of observations was much smaller in this exploratory analysis than in the main analyses, so future research should seek to confirm these exploratory findings. One way to achieve this would be to allow participants to practice the secondary task for a longer amount of time before any speech is presented, or to include a block with only the response time task and no speech to serve as a baseline condition that could enable the effects of improvement at the secondary task in the absence of speech to be partialled out. Another way to overcome the challenges introduced by practice effects is to employ a measure of listening effort in which practice effects are less likely to influence the results. To that end, Experiment 2 used the same speech materials as Experiment 1 but measured listening effort using pupillometry.

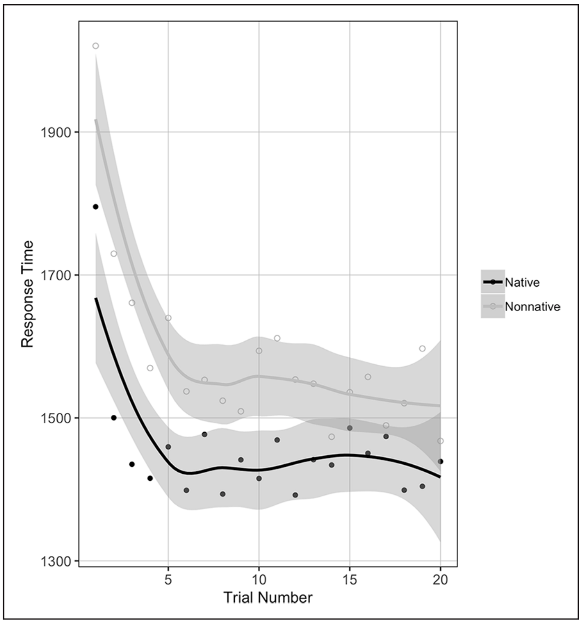

Figure 2. Scatterplot showing mean response times to the secondary visual response time task for the first 20 trials of the second block each participant completed. As in Figure 1, the lines represent loess curves, which were used for visualisation (and not analytical) purposes only.

Experiment 2

The pre-registration form for Experiment 2, which includes hypotheses, planned analyses, exclusion criteria, and sample size justification, can be found at https://osf.io/z74fp. We predicted that we would replicate the results of McLaughlin and Van Engen 2020 by showing that processing intelligible nonnative-accented speech results in greater pupil dilation, indicative of greater cognitive demand, than processing speech produced by a native talker. Further, we predicted that pupil dilation would reduce to a greater extent over the course of an experimental session (indicating greater reduction in listening effort) for nonnative- relative to native-accented speech.

Method

Participants. A total of 65 young adults (aged 18–23) participated in Experiment 2. Of these, 10 participants were excluded due to technical difficulties resulting in data loss (i.e., poor tracking of the pupil resulting in unusable data and/or computer malfunction), and 5 were excluded for failure to meet the eligibility criteria (i.e., they had previous exposure to Mandarin Chinese, did not speak English as a native language, and/or had familiarity with the target sentences from a prior experiment), leaving our pre-registered sample size of 50 participants for the analyses reported here. The study took approximately 1.5 hr to complete. As in Experiment 1, participants received course credit for participation, with the exception of four participants who received $15 for participation because the subject pool had closed.

Materials. The same sentences and talkers from Experiment 1 were used in Experiment 2. However, given that each trial was longer in this experiment than in Experiment 1, only 50 stimuli were used for each accent condition, resulting in 100 total sentences. This subset of the 160 sentences used in Experiment 1 was selected randomly after excluding sentences that we knew to be less intelligible than others.

Procedure. All data in the pupillometry task were collected with an EyeLink 1000 Plus camera in a sound-attenuating chamber, and pupil dilation was measured in arbitrary units (i.e., the units automatically provided by the EyeLink system, in this case representing the area of the pupil). The distances and settings for all equipment were based on recommendations from the EyeLink 1000 Plus manual. Data were collected at 500 Hz with a 35-mm lens in monocular mode, and participants rested their chin and forehead on a head mount during the task to reduce movement. The eye-tracking camera and computer monitor were placed approximately 21 in. and 30 in. from the head mount, respectively. The overhead lighting (set at a moderate brightness) and audio volume (set at a comfortable listening level) were consistent across participants.

The pupillometry task was preceded by detailed visual instructions. During each trial, a fixation cross was centred on a grey screen at all times. To limit data loss due to blinks, subjects were told to fixate on this cross throughout the trial and to reduce their blinking during periods in which the cross was red (as opposed to blue). The target stimulus was preceded by a quiet baseline period of 3,000 ms and was followed by a delay period of 3,000 ms; throughout this time window (on average 7,918 ms) the fixation cross was red. After the delay period, the fixation cross changed from red to blue, signalling to the participant that they should repeat the sentence aloud and that they could blink freely. Participants then pressed a foot pedal to continue to the next trial when they were ready. Before the onset of the next trial, the blue cross remained on-screen for a buffer period of 3,000 ms, allowing time for the phasic pupil response to stabilise. All timing was the same across conditions.

Breaks were inserted every 10 trials to reduce fatigue and task disengagement, which can cause pupil dilation to decrease over the course of an experiment (see, for example, McGarrigle et al., 2016). Given our assumption that decreases in pupil dilation reflect reductions in listening effort as a result of perceptual adaptation, it was especially important to control for potential effects of fatigue. As recommended by Winn and colleagues (2018), the research assistant conducting the experiment asked the participant brief questions that were unrelated to the task during these breaks, and participants were also required to sit back from the headrest. Before resuming trials after each break, the calibration of the eye-tracker was manually checked by the research assistant. The research assistant remained in the room during the task.

Stimuli were blocked by accent type, and the order of the accent blocks was counterbalanced across participants. In addition, the target sentences were counterbalanced such that half of the participants heard a given target in the native accent and the other half heard that target in the nonnative accent. The order of stimulus presentation within each block was randomised. Participants completed five practice trials before each block. Between blocks, participants were instructed to leave the sound-attenuating chamber for a 5-min break. After the pupillometry task, participants completed questionnaires regarding their demographic background and language experience. The experiment lasted approximately 1.5 hr in total.

Pupil data pre-processing. Pre-processing of the pupil data was done in R using the *gazeR* package following steps recommended by Geller and colleagues (2020). First, trials with more than 20% data loss due to blinks and participants with more than 20% trial loss in either accent condition were excluded. Our original pre-registered plan was to exclude trials with more than 50% data loss, but we decided to use a more conservative threshold based on recommendations for the *gazeR* package (Geller et al., 2020). Periods of missing data were identified as blinks. Due to noise created by the movement of the eyelid, all of these periods of missing data were extended such that the data 100 ms prior to and 200 ms after the blink were also removed. Linear interpolation was used to fill in the missing data. Next, a five-point moving average was used to smooth the data. The smoothed data were then normalised with subtractive baselining (Reilly et al., 2019), which is necessary for comparing across trials and conditions. The baseline for each trial was the median value of the 500 ms immediately preceding the onset of the stimulus. Finally, the data were time-binned to make model convergence less computationally demanding, essentially reducing the sampling frequency from 500 to 50 Hz. There was approximately 2% trial loss across participants in the native condition and approximately 5% trial loss across participants in the nonnative condition; this difference amounts to approximately 1.5 trials on average.

Results and discussion

Growth curve analysis (GCA; Mirman, 2014) was implemented using the *lme4* package (version 1.1.21) in R to model the shape of the pupil response, and *p* values for model parameters were obtained via the *lmerTest* package (version 3.0.1). GCA is a method of mixed effects modelling similar to polynomial regression that reduces collinearity by orthogonalizing the polynomial time terms. Fixed effects in the full model included the linear, quadratic, and cubic orthogonalized polynomial time terms (i.e., measures of time within a given trial), accent (i.e., native vs nonnative), trial number (i.e., a measure of time across the experimental session), and two- and three-way interactions between each of these main effects. The degree of the polynomial function was initially determined based on visual inspection of the data. The main effects of the linear ($\beta = 1{,}083.75$, $SE = 95.60$, $t = 11.34$, $p < .001$), quadratic ($\beta = -198.68$, $SE = 39.66$, $t = -5.01$, $p < .001$), and cubic ($\beta = -99.54$, $SE = 26.66$, $t = -3.73$, $p < .001$) polynomial terms were all significant, indicating that they contributed to the shape of the model fit for the pupil response curve. The main effect of accent was dummy-coded with the native condition as the reference level, and trial number was treated as a continuous predictor. The random effects structure included intercepts and slopes for participants, items, and the interaction between participants and accent (as recommended by Mirman, 2014).

A full summary of the selected model in Experiment 2 is included in Table S1 in the Supplementary Materials. Here, we limit our discussion to the main effects and interactions that directly pertain to our hypotheses. Consistent with our hypothesis, the significant main effect of accent indicated that on average, pupil dilation was greater for nonnative- relative to native-accented speech within the time window of interest ($\beta = 115.50$, $SE = 34.43$, $t = 3.36$, $p = .001$; Figure 3). Further, the significant interaction between accent and the linear polynomial time term indicated that pupil dilation increased more rapidly after sentence onset for nonnative- relative to native-accented speech ($\beta = 566.92$, $SE = 88.33$, $t = 6.42$, $p < .001$), consistent with previous research (McLaughlin & Van Engen, 2020).

The significant main effect of trial number indicated that pupil dilation decreased across trials for both accents ($\beta = -1.44$, $SE = 0.05$, $t = -27.57$, $p < .001$), likely reflecting both adaptation to the particular talker's voice and fatigue over time (McGarrigle et al., 2016; Winn et al., 2018). Most notably, the interaction between accent and trial number indicated that pupil dilation decreased more rapidly for the nonnative relative to the native accent across trials ($\beta = -2.29$, $SE = 0.07$, $t = -30.68$, $p < .001$; Figure 4). The full model had a marginal R^2 value of $R^2_{LMM(m)} = .05$ and a conditional R^2 value of $R^2_{LMM(c)} = .24$. These values were obtained via the *MuMIn* package (version 1.42.1).

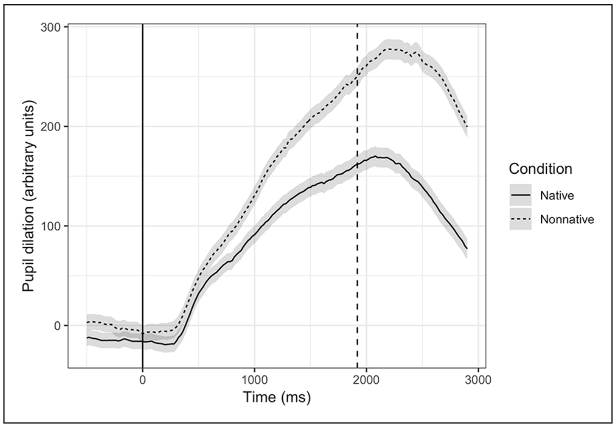

Figure 3. Pupil dilation over the course of a trial for native versus nonnative speech. Solid and dashed lines represent raw mean values, and grey ribbons represent one standard error. Sentence onset is denoted with a solid vertical line at 0 ms, and the average offset of all sentences is denoted with a dashed vertical line at 1,918 ms. This plot depicts the main effect of accent type, indicating that pupil dilation tends to be larger during nonnative- relative to native-accented speech processing.

As in Experiment 1, we opted to conduct an additional exploratory analysis on the data from Experiment 2 to assess whether the decrease in listening effort (as indexed by the decrease in pupil dilation) was primarily happening early in each block. We decided to test for an interaction between accent and trial number in the first 20 trials of each block by subsetting our data. We also assessed whether this interaction was present in the last 30 trials of each block in a separate model. The second exploratory model was built to allow us to compare the magnitude of the effect in each set of trials. That is, we aimed to determine whether the interaction was present for the first 20 trials, and whether the magnitude of the effect was stronger in the first 20 than in the last 30 trials. Note that unlike in the exploratory analysis in Experiment 1, the design for the present analysis remains within-subjects because we subsetted the data in both blocks.

Table S2 in the Supplemental Material summarises the models for the first 20 ("early") and last 30 ("late") trials. Most notably, the key interaction between accent and trial number was significant when analysing both the early ($\beta = -6.458$, $SE = 0.289$, $t = -22.384$, $p < .001$) and late ($\beta = -1.216$, $SE = 0.159$, $t = -7.640$, $p < .001$) trials, but the magnitude of the effect is considerably larger in the first 20 trials. This outcome reflects the pattern depicted in Figure 4: in approximately the first 20 trials, the pupil response in trials 1–10 versus 11–20 changes more in the nonnative condition than in the native condition, but this interaction is less pronounced in the later trials. Further, the variance explained by the model for the first 20 trials, $R^2_{LMM(m)} = .07$; $R^2_{LMM(c)} = .37$), was greater than that explained by the model for the last 30 trials, ($R^2_{LMM(m)} = .03$;

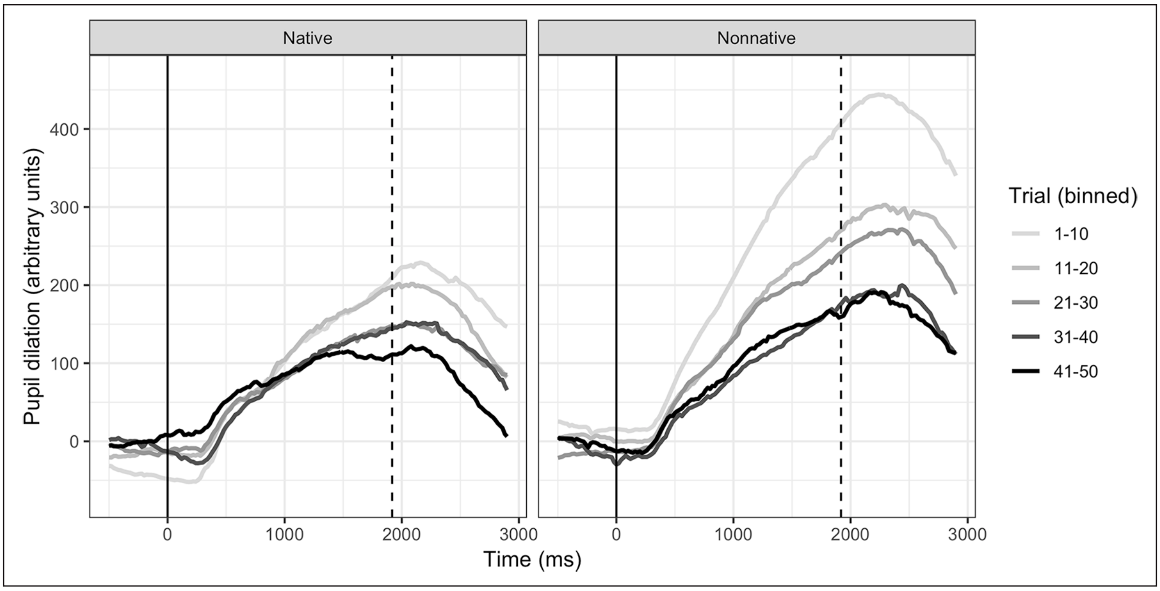

Figure 4. Pupil dilation over the course of a trial for native versus nonnative speech, grouped by bins of trials (see legend). Note that trials were binned for visual purposes only, and that the statistical model treated trial number as a continuous predictor. A solid vertical line at 0 ms and a dashed vertical line at 1,918 ms mark sentence onset and average offset, respectively. This plot depicts the interaction between accent type and trial number, indicating that the magnitude of the pupil response decreases more rapidly over the course of the experiment for nonnative- relative to native-accented speech.

$R^2_{LMM(c)} = .24$. These results suggest that much of the reduction in listening effort due to perceptual adaptation to the nonnative accent occurred within the early trials of the experiment.

Consistent with the findings of Experiment 1, the pupillometry data in Experiment 2 indicate that greater listening effort was imposed by nonnative- relative to native-accented speech. In addition, the interaction between accent and trial number suggests that perceptual adaptation to the nonnative accent reduces this additional listening effort over the course of a block to a greater extent than it does for native speech.

General discussion

Given concerns that multiple measures of listening effort may not be measuring the same underlying construct (Alhanbali et al., 2019; Strand et al., 2018), this study used both a behavioural and a physiological paradigm to assess listening effort. In this way, we aimed to ensure that any effects obtained were not specific to a single measure. The results from the two tasks were largely consistent with one another, providing converging evidence that (1) processing nonnative-accented speech requires greater listening effort than processing native-accented speech, as indicated by a main effect of accent, (2) the amount of listening effort required to process speech decreases rapidly with exposure, as indicated by a main effect of trial number, and (3) the reduction in listening effort over time may occur more rapidly for nonnative relative to native speech, as indicated by an interaction between accent and trial number in Experiment 2. Thus, it appears that although processing nonnative-accented speech requires greater listening effort than processing native-accented speech—even when intelligibility is equated across these conditions—these differences become less pronounced with exposure (replicating and extending the findings of McLaughlin & Van Engen, 2020). We did not explicitly test whether differences in effort between the native and nonnative conditions persisted following adaptation, but visual inspection of Figures 1 and 4 suggest that listeners still expended greater effort in the nonnative condition following adaptation.

The finding that listeners expended greater effort overall when processing fully intelligible nonnative- relative to native-accented speech is consistent with the general claims of the ELU model (Rönnberg et al., 2013), which postulates that in ideal listening conditions, phonological information is rapidly and automatically matched to lexical representations stored in memory. However, when there is a mismatch between the input and items in memory (e.g., due to background noise, reverberation, hearing impairment, or accented speech), working memory is recruited to support listening (Rönnberg et al., 2010). In the nonnative accent condition, in which the unfamiliar accent does not match stored representations of words in native listeners' lexicons, additional cognitive resources were required to process the speech, leaving fewer resources available to complete the secondary response time task (Experiment 1) or resulting in increased pupil dilation (Experiment 2).

The experiments reported here also provide the first evidence that listeners adapt to native and nonnative speech over time such that they expend less listening effort as they become more familiar with the accent or specific talker's voice, as indicated by the overall decrease in response times and pupil dilation over the course of the experiment for both talkers. These findings regarding listening effort are conceptually similar to previous work showing that familiarity with an accent or talker facilitates *identification* of speech (though see Drozdova et al., 2019, for evidence that this may not extend to nonnative listeners with low proficiency). Indeed, intelligibility of accented speech improves with exposure (Bradlow & Bent, 2008; Clarke & Garrett, 2004), even independent of the talker or the particular accent (Baese-Berk et al., 2013; Bradlow & Bent, 2008; Sidaras et al., 2009), and being familiar with the talker improves the listener's ability to recognise spoken words in background noise (Nygaard et al., 1994). However, this study is unique in its use of fully intelligible speech and its focus on listening effort rather than intelligibility; previous research has tended to use speech materials that vary in their intelligibility (though see Porretta & Tucker, 2019) and has typically focused on changes in intelligibility over time (Bradlow & Bent, 2008) or in the time it takes to process speech, which may be conflated with changes in intelligibility (Clarke & Garrett, 2004). Thus, these results indicate that adaptation to speech over time can reduce listening effort in addition to improving intelligibility.

Not only do listeners adapt to the talker such that listening effort is reduced with increased exposure, but this adaptation occurs quite rapidly, as it does for intelligibility (e.g., Bradlow & Bent, 2008). Indeed, as depicted in Figure 1 (which plots response times to the secondary visual task over the course of the experiment) and Figure 4 (which plots pupil dilation over the course of the experiment), the negative association between trial number and both response time and pupil dilation appears to become less pronounced for both talkers around the 20th trial (see also Clarke & Garrett, 2004). This is also indicated by the exploratory analysis of the Experiment 2 data, which shows that the magnitude of the decrease in pupil dilation is larger in the first 20 than the last 30 trials, particularly for the nonnative accent. Although the decrease in response times across trials in Experiment 1 is likely to be partially attributable to practice effects (i.e., response times decrease over the course of the experiment because participants improve at the secondary task), these results also suggest that individuals adapt to the speech of both native and nonnative unfamiliar talkers. This claim is supported by the Experiment 1 exploratory analysis, which was conducted on the second block only after participants had ample practice with the task in the first block. In that analysis, response times appeared to plateau much earlier than the 20th trial, suggesting that the adaptation effect occurs even more rapidly than would be implied by examining changes in response times over the first 20 trials of the first block, in which adaptation and practice effects are confounded.

This work provides the first demonstration that adaptation as measured by changes in listening effort occurs more rapidly for nonnative- relative to native-accented speech, at least for the two talkers we used here. Indeed, processing nonnative-accented speech is initially more cognitively demanding than processing native-accented speech, so there is greater room for reductions in listening effort, leading to stronger effects of adaptation. The mechanism for adaptation remains to be determined, but listeners may be either adjusting their phoneme categories (i.e., perceptual learning; Norris et al., 2003) or relaxing their phoneme category boundaries when listening to nonnative-accented speech (Zheng & Samuel, 2019). These effects emerged to some extent in both experiments, but the pupillometry data provided stronger evidence than the behavioural data that listeners adapt to nonnative-accented speech more rapidly than native-accented speech. One explanation for the lack of an accent-by-trial number interaction in the pre-registered analysis in Experiment 1 may be that the effect is subtle and the dual-task paradigm we used was not sufficiently sensitive to detect the interaction. Measures of listening effort differ markedly in their ability to detect changes in listening effort that arise from changes in the level of the background noise in a speech perception task (Strand et al., 2018), so it is possible that this behavioural measure was simply not sensitive enough to detect changes in the listening effort associated with processing nonnative- versus native-accented speech over time. Alternatively, it is also possible that practice effects in the complex dual-task mask the accent-by-trial number interaction—that is, participants may continually improve at the response time task for the first 20 or so trials, and if those effects are sufficiently large and adaptation happens sufficiently quickly, this may obscure any differences in adaptation across accents.

The significant interaction between accent and trial number in Experiment 2 indicates that reductions in pupil dilation occur more quickly for nonnative relative to native speech. This is in line with our hypothesis that listeners would show greater adaptation to the nonnative relative to the native accent. Alternatively, it is also possible that these data reflect the fact that listeners became fatigued more quickly in the nonnative accent condition. Prior research has shown that difficult listening conditions elicit greater fatigue than easy conditions, resulting in reduced physiological arousal as indicated by decreases in pupil responses over time (see McGarrigle et al., 2016). However, this explanation for the observed interaction in Experiment 2 seems unlikely for several reasons. First, the increased fatigue in difficult listening conditions in the study by McGarrigle and colleagues (2016) was indicated by a steeper decline after the peak pupil diameter in a difficult relative to an easy level of background noise. In this study, however, pupil dilation appears to decline at comparable rates following the peak in the native and nonnative conditions (see Figure 3). Second, the stimuli in this study were substantially shorter than those in McGarrigle et al.'s (2016) study (approximately 2 s vs 13–18 s), so participants spent less time on every trial actively processing speech, reducing the opportunity for fatigue to occur and influence the pupil response. Third, our stimuli were presented without background noise and intelligibility was above 96% across accents, so it is unlikely that our participants experienced fatigue to the same extent as those in McGarrigle et al. (2016). Finally, in an attempt to minimise fatigue effects, participants in Experiment 2 took short breaks in which they removed their heads from the headrest and engaged in short conversations with the experimenter (as suggested by Winn et al., 2018) every 10 trials. Despite these efforts, it is possible that fatigue may still may have occurred and influenced pupillary responses, and future research on adaptation using pupillometry should systematically address this possibility.

This study adds to a growing body of research demonstrating a dissociation between the intelligibility of speech and the effort necessary to process it. For example, noise-reduction algorithms (Sarampalis et al., 2009) and task demands (Mackersie & Cones, 2011) can affect listening effort even in situations in which speech intelligibility is equivalent across conditions. In this study, we controlled for intelligibility differences between the native and nonnative accents by using highly intelligible sentences and only analysing trials in which all four keywords were identified correctly. These findings suggest that clinical evaluations that rely exclusively on measures of intelligibility may miss important information about the listener's experience.

One limitation of the current research is the reliance on a single native and nonnative talker. Although we demonstrated differences in listening effort between these two talkers, future research should use a greater number of accents and talkers to ensure that our findings are not specific to the individual talkers we used in these experiments. Nonnative-accented talkers differ in their intelligibility (e.g., Bradlow & Bent, 2008) and may also differ in the extent to which their speech incurs additional listening effort (see Van Engen & Peelle, 2014). That is, speech that is more unfamiliar or results in a greater mismatch between the acoustic input and representations stored in memory is likely to require greater recruitment of cognitive resources, and talkers differ in the extent to which their speech elicits phonological mismatches in a listener's lexicon. Thus, future research should seek to evaluate the generalisability of the findings reported here with a larger range of accents and talkers, or with a single speaker recorded in both their native accent and in a constructed, artificial accent (e.g., Banks et al., 2015; Janse & Adank, 2012). The latter approach has the benefit of removing any speaker-related confounds, but may also have limited ecological validity. That is, constructed accents are often made by asking speakers to read an orthographic transcription in which phonemes are systematically changed, and though the constructed accent using this method should contain segmental deviations similar to those present in a natural accent, these artificial accents may lack the suprasegmental qualities of a natural accent, such as prosodic deviations.

Future research should also seek to test the robustness of these findings across listener experiences. These experiments deliberately included participants with no familiarity with the nonnative accent in an attempt to make the effort required for the native and nonnative conditions as dissimilar as possible. However, it is not clear whether these effects would persist for listeners who have greater familiarity with the accent included in the experiment. Porretta and Tucker (2019) showed that individuals with greater familiarity with a given accent expend less listening effort when processing speech in that accent relative to individuals who are less familiar with the accent. Thus, future work should evaluate whether greater previous experience with an accent additionally increases the extent to which listeners adapt to that accent as measured by reductions in listening effort over time. Our experiments used listeners with little to no experience with Mandarin Chinese–accented English, but it is possible that individuals with greater familiarity with this accent would not show effort differences between native- and nonnative-accented speech, or may adapt to nonnative speech even more quickly than the individuals in this study.

Acknowledgements

The authors thank the undergraduate research assistants at Washington University in St. Louis for assistance with data collection.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by Washington University in St. Louis, and by the National Science Foundation through Graduate Research Fellowships awarded to Violet Brown and Drew McLaughlin (DGE-1745038).

ORCID iD

Violet A Brown https://orcid.org/0000-0001-5310-6499

Open practices

The data and materials from the present experiment are publicly available at the Open Science Framework website: https://osf.io/xajdw/

Supplemental material

The supplemental material is available at qjep.sagepub.com

References

Alexander, J. E. D., & Nygaard, L. C. (2019). Specificity and generalization in perceptual adaptation to accented speech. *The Journal of the Acoustical Society of America*, *145*(6), 3382.

Alhanbali, S., Dawes, P., Millman, R. E., & Munro, K. J. (2019). Measures of listening effort are multidimensional. *Ear and Hearing*, *40*, 1084–1097. https://doi.org/10.1097/AUD.0000000000000697

Alsius, A., Navarra, J., & Soto-Faraco, S. (2007). Attention to touch weakens audiovisual speech integration. *Experimental Brain Research*, *183*(3), 399–404.

Baese-Berk, M. M., Bradlow, A. R., & Wright, B. A. (2013). Accent-independent adaptation to foreign accented speech. *The Journal of the Acoustical Society of America*, *133*(3), EL174–EL180.

Banks, B., Gowen, E., Munro, K. J., & Adank, P. (2015). Cognitive predictors of perceptual adaptation to accented speech. *The Journal of the Acoustical Society of America*, *137*(4), 2015–2024.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Bartoń, K. (2019). *Package "MuMIn" (Version 1.43.6)*. https://cran.r-project.org/web/packages/MuMIn/MuMIn.pdf

Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R., Singmann, H., Dai, B., Scheipl, F., Grothendieck, G., & Green, P. (2014). *Package "lme4" (Version 1.1-15). R foundation for statistical computing, Vienna*, *12*. https://github.com/lme4/lme4/

Beatty, J. (1982). Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological Bulletin*, *91*(2), 276–292.

Borghini, G., & Hazan, V. (2018). Listening effort during sentence processing is increased for non-native listeners: A pupillometry study. *Frontiers in Neuroscience*, *12*, e152.

Bourland-Hicks, C., & Tharpe, A. M. (2002). Listening effort and fatigue in school-age children with and without hearing loss. *Journal of Speech, Language, and Hearing Research: JSLHR*, *45*(3), 573–584.

Bradlow, A. R., & Bent, T. (2008). Perceptual adaptation to non-native speech. *Cognition*, *106*(2), 707–729.

Brown, V. A., & Strand, J. F. (2018). Noise increases listening effort in normal-hearing young adults, regardless of working memory capacity. *Language, Cognition and Neuroscience*, *34*, 628–640.

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.

Desjardins, J. L., & Doherty, K. A. (2013). Age-related changes in listening effort for various types of masker noises. *Ear and Hearing*, *34*(3), 261–272.

Desjardins, J. L., & Doherty, K. A. (2014). The effect of hearing aid noise reduction on listening effort in hearing-impaired adults. *Ear and Hearing*, *35*(6), 600–610.

Downs, D. W. (1982). Effects of hearing aid use on speech discrimination and listening effort. *The Journal of Speech and Hearing Disorders*, *47*(2), 189–193.

Downs, D. W., & Crum, M. A. (1978). Processing demands during auditory learning under degraded listening conditions. *Journal of Speech and Hearing Research*, *21*(4), 702–714.

Drozdova, P., van Hout, R., & Scharenborg, O. (2019). Talker-familiarity benefit in non-native recognition memory and word identification: The role of listening conditions and proficiency. *Attention, Perception & Psychophysics*, *81*(5), 1675–1697.

Fraser, S., Gagné, J.-P., Alepins, M., & Dubois, P. (2010). Evaluating the effort expended to understand speech in noise using a dual-task paradigm: The effects of providing visual speech cues. *Journal of Speech, Language, and Hearing Research: JSLHR*, *53*(1), 18–33.

Gagné, J.-P., Besser, J., & Lemke, U. (2017). Behavioral assessment of listening effort using a dual-task paradigm: A review. *Trends in Hearing*, *21*, 2331216516687287.

Geller, J., Winn, M. B., Mahr, T., & Mirman, D. (2020). GazeR: A package for processing gaze position and pupil size data. *Behavioral Research Methods*. Advance online publication. https://doi.org/10.3758/s13428-020-01374-8

Janse, E., & Adank, P. (2012). Predicting foreign-accent adaptation in older adults. *Quarterly Journal of Experimental Psychology*, *65*(8), 1563–1585.

Johnson, J., Xu, J., Cox, R., & Pendergraft, P. (2015). A comparison of two methods for measuring listening effort as part of an audiologic test battery. *American Journal of Audiology*, *24*(3), 419–431.

Kahneman, D. (1973). *Attention and effort*. PrenticeHall.

Klingner, J., Tversky, B., & Hanrahan, P. (2011). Effects of visual and verbal presentation on cognitive load in vigilance, memory, and arithmetic tasks. *Psychophysiology*, *48*(3), 323–332.

Kraljic, T., & Samuel, A. G. (2005). Perceptual learning for speech: Is there a return to normal? *Cognitive Psychology*, *51*(2), 141–178.

Kuchinsky, S. E., Ahlstrom, J. B., Vaden, K. I., Jr., Cute, S. L., Humes, L. E., Dubno, J. R., & Eckert, M. A. (2013). Pupil size varies with word listening and response selection difficulty in older adults with hearing loss. *Psychophysiology*, *50*(1), 23–34.

Kuznetsova, A., Brockhoff, P., & Christensen, R. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1–26.

Mackersie, C. L., & Cones, H. (2011). Subjective and psychophysiological indexes of listening effort in a competing-talker task. *Journal of the American Academy of Audiology*, *22*(2), 113–122.

Mattys, S. L., Davis, M. H., Bradlow, A. R., & Scott, S. K. (2012). Speech recognition in adverse conditions: A review. *Language and Cognitive Processes*, *27*, 953–978.

McCoy, S. L., Tun, P. A., Cox, L. C., Colangelo, M., Stewart, R. A., & Wingfield, A. (2005). Hearing loss and perceptual effort: Downstream effects on older adults' memory for speech. *The Quarterly Journal of Experimental Psychology, Section A: Human Experimental Psychology*, *58*(1), 22–33.

McGarrigle, R., Dawes, P., Stewart, A. J., Kuchinsky, S. E., & Munro, K. J. (2016). Pupillometry reveals changes in physiological arousal during a sustained listening task. *Psychophysiology*, *54*(2), 193–203.

McLaughlin, D. J., & Van Engen, K. J. (2020). Task-evoked pupillary response for intelligible accented speech. *The Journal of the Acoustical Society of America*, *147*, EL151. https://doi.org/10.1121/10.0000718

Mirman, D. (2014). *Growth curve analysis and visualization using R (Chapman & Hall/CRC The R Series)* (1st ed.). Chapman and Hall/CRC.

Nakagawa, S., Johnson, P. C. D., & Schielzeth, H. (2017). The coefficient of determination R^2 and intra-class correlation coefficient from generalized linear mixed-effects models revisited and expanded. *Journal of the Royal Society, Interface*, *14*(134), 1–11. https://doi.org/10.1098/rsif.2017.0213

Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R^2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, *4*(2), 133–142.

Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology*, *47*(2), 204–238.

Nygaard, L. C., Sommers, M. S., & Pisoni, D. B. (1994). Speech perception as a talker-contingent process. *Psychological Science*, *5*(1), 42–46.

Peelle, J. E. (2018). Listening effort: How the cognitive consequences of acoustic challenge are reflected in brain and behavior. *Ear and Hearing*, *39*, 204–214. https://doi.org/10.1097/AUD.0000000000000494

Pichora-Fuller, M. K., Kramer, S. E., Eckert, M. A., Edwards, B., Hornsby, B. W. Y., Humes, L. E., Lemke, U., Lunner, T., Matthen, M., Mackersie, C. L., Naylor, G., Phillips, N. A., Richter, M., Rudner, M., Sommers, M. S., Tremblay, K. L., & Wingfield, A. (2016). Hearing impairment and cognitive energy: The Framework for Understanding Effortful Listening (FUEL). *Ear and Hearing*, *37*(Suppl. 1), 5S–27S.

Picou, E. M., & Ricketts, T. A. (2014). The effect of changing the secondary task in dual-task paradigms for measuring listening effort. *Ear and Hearing*, *35*(6), 611–622.

Porretta, V., & Tucker, B. V. (2019). Eyes wide open: Pupillary response to a foreign accent varying in intelligibility. *Frontiers in Communication*, *4*, Article 8.

R Core Team. (2016). *R: A language and environment for statistical computing. R Foundation for statistical computing*. http://www.R-project.org/

Reilly, J., Kelly, A., Kim, S. H., Jett, S., & Zuckerman, B. (2019). The human task-evoked pupillary response function is linear: Implications for baseline response scaling in pupillometry. *Behavior Research Methods*, *51*(2), 865–878.

Rönnberg, J., Holmer, E., & Rudner, M. (2019). Cognitive hearing science and ease of language understanding. *International Journal of Audiology*, *58*(5), 247–261.

Rönnberg, J., Lunner, T., Zekveld, A., Sörqvist, P., Danielsson, H., Lyxell, B., Dahlström, O., Signoret, C., Stenfelt, S., Pichora-Fuller, M. K., & Rudner, M. (2013). The Ease of Language Understanding (ELU) model: Theoretical, empirical, and clinical advances. *Frontiers in Systems Neuroscience*, *7*, Article 31.

Rönnberg, J., Rudner, M., Foo, C., & Lunner, T. (2008). Cognition counts: A working memory system for ease of language understanding (ELU). *International Journal of Audiology*, *47*(Suppl. 2), S99–S105.

Rönnberg, J., Rudner, M., Lunner, T., & Zekveld, A. A. (2010). When cognition kicks in: Working memory and speech understanding in noise. *Noise & Health*, *12*(49), 263–269.

Rudner, M., Ng, E. H. N., Ronnberg, N., Mishra, S., Ronnberg, J., Lunner, T., & Stenfelt, S. (2011). Cognitive spare capacity as a measure of listening effort. *Journal of Hearing Science*, *1*(2), 47–49. http://www.diva-portal.org/smash/record.jsf?pid=diva2%3A792812&dswid=640

Sarampalis, A., Kalluri, S., Edwards, B., & Hafter, E. (2009). Objective measures of listening effort: Effects of background noise and noise reduction. *Journal of Speech, Language, and Hearing Research: JSLHR*, *52*(5), 1230–1240.

Seeman, S., & Sims, R. (2015). Comparison of psychophysiological and dual-task measures of listening effort. *Journal of Speech, Language, and Hearing Research: JSLHR*, *58*(6), 1781–1792.

Sidaras, S. K., Alexander, J. E. D., & Nygaard, L. C. (2009). Perceptual learning of systematic variation in Spanish-accented speech. *The Journal of the Acoustical Society of America*, *125*(5), 3306–3316.

Strand, J. F., Brown, V. A., Merchant, M. B., Brown, H. E., & Smith, J. (2018). Measuring listening effort: Convergent validity, sensitivity, and links with cognitive and personality measures. *Journal of Speech, Language, and Hearing Research: JSLHR*, *61*, 1463–1486.

Van Engen, K. J., Chandrasekaran, B., & Smiljanic, R. (2012). Effects of speech clarity on recognition memory for spoken sentences. *PLOS ONE*, *7*(9), e43753.

Van Engen, K. J., & McLaughlin, D. J. (2018). Eyes and ears: Using eye tracking and pupillometry to understand challenges to speech recognition. *Hearing Research*, *369*, 56–66. https://doi.org/10.1016/j.heares.2018.04.013

Van Engen, K. J., & Peelle, J. E. (2014). Listening effort and accented speech. *Frontiers in Human Neuroscience*, *8*, e577.

Winn, M. B., Edwards, J. R., & Litovsky, R. Y. (2015). The impact of auditory spectral resolution on listening effort revealed by pupil dilation. *Ear and Hearing*, *36*(4), e153–e165.

Winn, M. B., Wendt, D., Koelewijn, T., & Kuchinsky, S. E. (2018). Best practices and advice for using pupillometry to measure listening effort: An introduction for those who want to get started. *Trends in Hearing*, *22*, 2331216518800869.

Zekveld, A. A., Kramer, S. E., & Festen, J. M. (2010). Pupil response as an indication of effortful listening: The influence of sentence intelligibility. *Ear and Hearing*, *31*(4), 480–490.

Zheng, Y., & Samuel, A. G. (2019). The relationship between phonemic category boundary changes and perceptual adjustments to natural accents. *Journal of Experimental Psychology. Learning, Memory, and Cognition*. Advance online publication. https://doi.org/10.1037/xlm0000788