Sequence Effects and Speech Perception: Cognitive Load for Speaker-Switching Within and Across Accents

Drew J. McLaughlin

Jackson S. Colvett

Julie M. Bugg

Kristin J. Van Engen

Department of Psychological & Brain Sciences, Washington University in St. Louis

Author Note

Acknowledgements: This work was supported by National Science Foundation Graduate Research Fellowship DGE-1745038 (Drew J. McLaughlin). Thank you to Jaida Dalton and Lindsay Nobles for assistance with data collection.

The authors have no relevant financial or non-financial conflicts of interest to disclose.

Correspondence concerning this article should be addressed to Drew J. McLaughlin, Department of Psychological & Brain Sciences, Washington University in Saint Louis, One Brookings Dr, St. Louis, MO 63130. Email: drewjmclaughlin@wustl.edu. Phone: 314-935-2744

Abstract

Prior work in speech perception indicates that listening tasks with multiple speakers (as opposed to a single speaker) result in slower and less accurate processing. Notably, the trial-to-trial cognitive demands of switching between speakers or switching between accents have yet to be examined. We used pupillometry, a physiological index of cognitive load, to examine the demands of processing native- and nonnative-accented speech when listening to sentences produced by the same speaker consecutively (no switch), a novel speaker of the same accent (within-accent switch), and a novel speaker with a different accent (across-accent switch). Inspired by research on sequential adjustments in cognitive control, we anticipated that by examining the trial-to-trial demands of speech perception we would be able to identify the cognitive demands of accommodating a novel speaker and accent. Our results indicate that switching between speakers was more cognitively demanding than listening to the same speaker consecutively. Additionally, switching to a novel speaker with a different accent was more cognitively demanding than switching between speakers of the same accent. However, there was an asymmetry for across-accent switches, such that switching from native to nonnative accent was more demanding than the reverse. Findings from the present study align with work examining multi-talker processing costs and provide novel evidence that listeners dynamically adjust cognitive resources to accommodate speaker and accent variability. We discuss these novel findings in the context of congruency sequence effects and a linguistic account of the cognitive demands of resolving acoustic-phonetic variability.

Keywords: pupillometry, speech perception, congruency sequence effect, accent

Sequence Effects and Speech Perception: Cognitive Load for Speaker-Switching Within and Across Accents

Despite substantial acoustic variation in how speakers produce speech, listeners can understand speech seemingly effortlessly. However, the challenge that listeners face when mapping acoustic input onto their linguistic representations can be complicated by individual speaker variability and accent variability. For example, in word recognition experiments responses are typically slower and less accurate in blocks with multiple speakers (changing from trial-to-trial), as compared to blocks with a single speaker (Carter et al., 2019; Choi & Perrachione, 2019; Heald & Nusbaum, 2014; Magnuson et al., 2021; Mullennix et al., 1989; Saltzman et al., 2021; Wong et al., 2004). These *multi-talker processing costs* suggest that adapting to novel speakers is a cognitively demanding process.

This literature has yet to examine the processing costs associated with alternating between speakers with different accents. Research examining perception of nonnative ("foreign") accents has demonstrated that even highly fluent and intelligible nonnative speakers can be more cognitively demanding to understand than native speakers (Brown et al., 2020; McLaughlin & Van Engen, 2020). Systematic and idiosyncratic deviations in how speech is produced by nonnative speakers (as compared to native speakers) may also exacerbate perceptual demands in a multi-talker setting. By examining multi-talker processing costs in a multi-accent setting, one can determine whether the 'phonetic distance' between speakers affects the cognitive demands of the speaker accommodation process.

In the present study, we investigated whether the cognitive demands for switching between native speakers are similar to those for switching between nonnative speakers. Additionally, we assessed whether switching between speakers of different accents poses a greater processing cost than switching between speakers of the same accent. Our examination of multi-talker processing costs takes a novel approach – inspecting trial-to-trial changes in cognitive processing load with pupillometry.

Pupillometry

Pupillometry, the measure of pupil diameter over time, has been used across multiple domains as a physiological index of cognitive processing load (Beatty, 1982). By tracking the "task-evoked" pupil response, one can compare the cognitive demands imposed by different tasks or experimental manipulations. In speech perception, cognitive pupillometry has been applied widely (for a review, see Van Engen & McLaughlin, 2018), demonstrating a systematic relationship between the magnitude of the pupil response and the difficulty of speech perception. Pupil response increases as speech intelligibility decreases due to background noise (Zekveld et al., 2010; Zekveld & Kramer, 2014) or nonnative accent (Porretta & Tucker, 2019). For highly-intelligible materials (e.g., sentences that are fully understood by the listener), pupillometry has been used to reveal that increasing signal degradation results in larger pupil response (Winn et al., 2015), as does nonnative, as compared to native, accent (Brown et al., 2020; McLaughlin & Van Engen, 2020).

Pupillometry has also been used as a measure of cognitive demands in cognitive control tasks that require updating, switching, or inhibition (e.g., van der Wel & van Steenbergen, 2018). Moreover, pupillometry has been used to assess trial-to-trial sequence effects in cognitive control (van Steenbergen & Band, 2013). By examining the pupil response in the current trial (N) with regard to the condition of the previous trial (N-1), we can determine how preceding context affects current cognitive demands.

Sequence effects

While analyzing the effect of trial N-1 on trial N is a novel approach to assessing the demands of speech perception, it is used commonly in other cognitive science literatures. For example, the congruency sequence effect (CSE; Gratton et al., 1992) refers to a reduction in the performance difference between congruent (e.g., RED in red-colored ink) and incongruent (e.g., RED in blue-colored ink) trials in conflict tasks such as Stroop when the previous trial is incongruent as opposed to congruent (for reviews, see Duthoo et al., 2014a; Egner, 2007). The CSE is interpreted as an adaptive adjustment of control based on the previous trial type, such that people up-regulate control when trial N-1 is incongruent and are thus less susceptible to conflict on trial N (Botvinick et al., 2001; see Schmidt & Weissman, 2014, for an alternative account).

While the current study does not index adjustments based on the *congruency* of the previous trial, there are clear connections to the present linguistic design. The more difficult (nonnative) accent in our design is analogous to an incongruent Stroop trial. An upregulation of effort during a listening trial with a nonnative speaker may reduce the difficulty of speech perception on the subsequent trial. Additionally, it is possible that the multi-talker processing costs commonly observed in blocked experiments reflect listeners actively adjusting their perceptual mappings every trial (i.e., “accommodating” a speaker or accent). If this is the case, we should be able to examine these sequence effects directly.

Research questions and hypotheses

We investigated speaker sequence effects using pupillometry to assess three key research questions. First, is there a cost for switching between speakers (as opposed to listening to the same speaker)? Second, does the magnitude of a switching cost depend on the ‘phonetic distance’ between two speakers? That is, is switching between speakers with the same accent easier than switching between speakers with different accents? Finally, are all across-accent switches equally difficult? That is, will switching from a native accent to a nonnative accent be equivalent to switching from a nonnative accent to a native accent?

We report two experiments that serve as an initial test of how the speaker and accent on the previous trial affect listening effort on the current trial. We predicted that:

1. Cognitive load (as indexed by pupil diameter) would be greater when switching speakers than when repeating the same speaker.

2. Switching across accents would be more cognitively demanding than switching within an accent.

3. There would be an interaction, such that native-to-nonnative switches would be most difficult. This third hypothesis parallels the finding in the CSE literature where performance is worst for challenging incongruent trials preceded by the less challenging congruent trials.

Experiment 1

In Experiment 1, we re-analyzed data from McLaughlin & Van Engen (2020) to examine the effects of switching between a native and a nonnative speaker of English. Full methodological details can be found in the original paper. We summarize key points below.

Method

Pre-registration, materials, experiment, data, and analysis code for McLaughlin and Van Engen (2020) are available from https://osf.io/7dajv/. The current re-analysis of this data was not pre-registered. Data and analysis code are available from https://osf.io/ajmqz.

Dataset description

McLaughlin and Van Engen (2020) contains a sample of 52 young adult subjects (39 female and 13 male; Age *M* = 19.46, *SD* = 1.07) recruited from the Washington University

Psychology Subjects Pool. All subjects were screened for normal hearing and were native speakers of American English with little exposure to Mandarin Chinese.

Subjects' pupil response was tracked during presentation of sentence-length materials. Two speakers were presented during the session: a native American-accented speaker of English, and a nonnative Mandarin Chinese-accented speaker of English. Sixty trials were presented (30 per accent) in a randomized order. After each trial, subjects repeated the sentence aloud. Every three trials, using a scale of one to nine, subjects pressed a key to indicate how effortful it was to understand the previous speaker.

Subjects' responses were scored for recognition accuracy. Any trials in which keywords were missed were excluded from the dataset. Data were pre-processed following standard pupillometry procedures: blinks were identified, expanded, and extrapolated across; data were smoothed with a 10 Hz moving average window; data were baselined using the 500 ms of data immediately preceding stimulus onset (i.e., baseline values were subtracted from all values in the respective trial); and data were time-binned, reducing the sampling frequency from 500 Hz to 50 Hz. Trials with more than 50% missing data were excluded from analyses.

Preparation of dataset for novel analyses

A switch condition was added to the dataset by comparing the current trial's (N) accent condition against the previous trial's (N-1) accent condition. If the two trials matched, they were labeled "no switch", and if they did not match, they were labeled "switch". Trial 1 data was removed (as there was no preceding context), as were any trials following an excluded trial (i.e., due to blinks or intelligibility). A total of 80 trials (approximately 2.7%) were removed from the original dataset in this process.

Growth curve analysis

Growth curve analysis (GCA) was implemented with the lme4 R package (Bates et al., 2014) to examine the data. GCA is a mixed-effects modeling approach similar to polynomial regression (Mirman, 2016). Orthogonalized polynomial predictors (linear, quadratic, cubic, etc.) are incorporated into the fixed and random effects of the model, allowing for a time-course analysis that is non-linear. This analysis approach is frequently used for analyzing pupillometry data because the curve of a task-evoked pupil response can be fit with a polynomial basis (i.e., is often similar to a cubic shape). In GCA, fixed effects of conditions determine whether there are differences in overall magnitude between levels (i.e., shifting the curve vertically), and interactions between these fixed effects and the fixed effects of the polynomial parameters determine whether the shape of the pupil response differs by condition (i.e., does the rate of increase in pupil size differ by condition?). The random effect structure of all models included random intercepts for subjects and items, and random slopes of the linear, quadratic, and cubic polynomials nested within subjects and items.

Results

Table 1 summarizes all log-likelihood model comparisons from the growth curve analysis. The linear, quadratic, and cubic polynomials all significantly improved fit (all *p*'s < .001).

Table 1

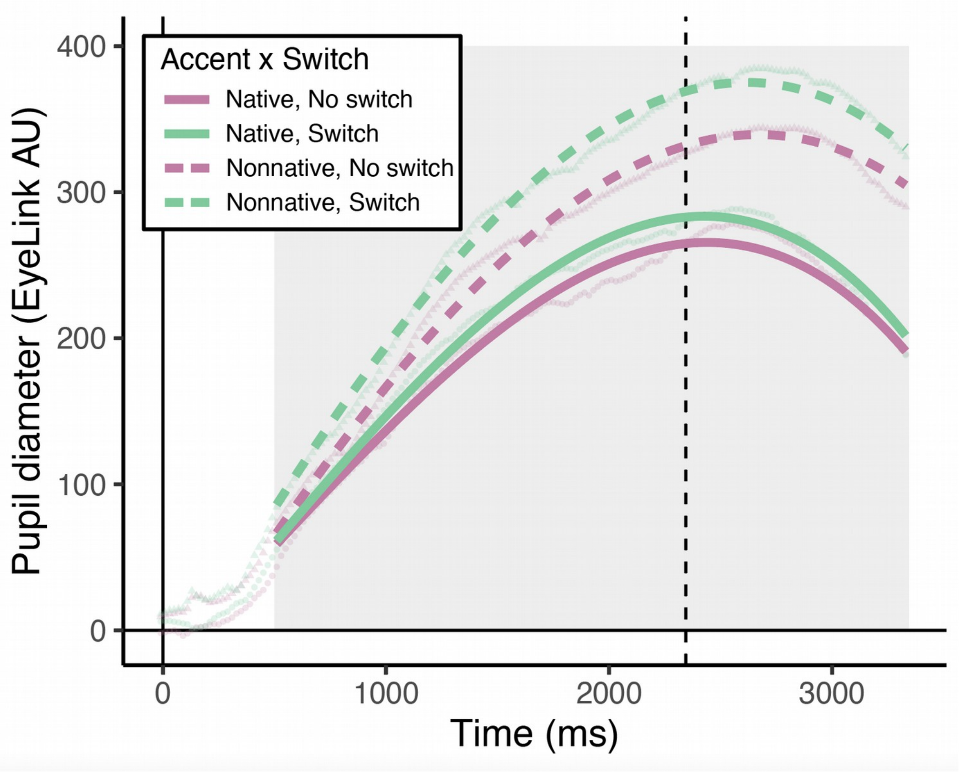
Log-likelihood Model Comparisons for Growth Curve Analysis of Experiment 1

Effect	χ^2	df	*p*
Linear polynomial	12609	1	< .001 ***
Quadratic polynomial	4790.60	1	< .001 ***
Cubic polynomial	74.37	1	< .001 ***
Accent (Levels: Native, Nonnative)	10.57	1	.001 **
Switch (Levels: No Switch, Switch)	303.43	1	< .001 ***
Accent x Switch	72.94	1	< .001 ***
Accent x Linear polynomial	16.54	1	< .001 ***
Accent x Quadratic polynomial	0.11	1	.74
Accent x Cubic polynomial	1.48	1	.22
Switch x Linear polynomial	0.27	1	.60
Switch x Quadratic polynomial	13.78	1	< .001 ***
Switch x Cubic polynomial	0	1	> .99

The fixed effects of accent (reference level: native) and switch (reference level: no switch) were both dummy-coded. Both effects significantly improved model fit (*p*’s ≤ .001). The direction of the accent estimate indicated that the nonnative accent condition elicited relatively larger pupil response than the native accent condition ($\boldsymbol{\beta}$ = 41.97, *p* < .001). For the effect of switch, the model estimate indicated a larger pupil response for speaker switches as compared to speaker repeats ($\boldsymbol{\beta}$ = 18.76, *p* < .001). Notably, the interaction between accent and switch also significantly improved model fit (*p* < .001), indicating that the effect of switching speakers was larger for the nonnative accent condition (Figure 1). Switching from the native to the nonnative speaker was costlier than repeating the same nonnative speaker, and costlier than switching from a nonnative to a native speaker. Interactions of the effects of accent and switch with the polynomial time terms are reported in Table 1.

Figure 1

Interaction of Accent and Switch Effects in Experiment 1



Note. The Experiment 1 interaction between accent and switch is shown with model fits (lines) and raw data means (points). Y-axis shows pupil diameter in EyeLink AU (Arbitrary Units), where zero is the baseline calculated to align data across trials. X-axis shows time in ms, beginning at trial start (zero). The dashed vertical line indicates the average offset time for all stimuli. The gray box indicates the window of the data used in analyses.

Discussion

The results of Experiment 1 suggest that switching between speakers is more cognitively demanding than listening to the same speaker consecutively. Additionally, there is an asymmetry in the switching costs, where switching from a native speaker to a nonnative speaker is particularly costly. While these results provide initial evidence of dynamic trial-to-trial processing adjustments to speakers and accents, they are limited by design choices made by McLaughlin and Van Engen (2020).

Experiment 2

In Experiment 2, we aimed to replicate and extend the findings from Experiment 1, but with three key changes to strengthen our design. First, we eliminated the subjective effort ratings, which were not relevant to the current aims.[1] Second, we increased the number of speakers in Experiment 2 (two native, two nonnative). This allowed us to examine switching costs when switching between two speakers of the same accent (in addition to switching across accents). Third, rather than randomly intermixing speakers, we designed our lists with an equal number of trials with no switches, within-accent switches, and across-accent switches in each block.

Our predictions are consistent with Experiment 1: speaker switches should be more difficult than speaker repetitions, across-accent switches should be more difficult than within-accent switches, and an asymmetry should emerge such that switching from native to nonnative accent is more demanding than switching from nonnative to native accent. Additionally, we predicted a conceptual replication of McLaughlin and Van Engen (2020), demonstrating that highly-intelligible nonnative-accented speech is more cognitively demanding than native-accented speech.

Method

Pre-registration, materials, experiment, data, and analysis code are all available from https://osf.io/ajmqz. The recruitment plan and protocol for this experiment was approved by the Washington University in St. Louis Institutional Review Board.

[1] The analysis in Experiment 1 assumed that any interference of this measure (which was acquired every three trials) with the sequence effects data was equally spread across conditions. By removing this feature entirely from Experiment 2, we were able to confirm this assumption and remove a potential source of noise from the task.

Participants

Sixty-three young adult participants (46 female, 17 male; Age *M* = 19.68, *SD* = 1.10) were recruited from the Washington University Psychology Subject Pool through SONA Systems. Recruitment for the study began before the COVID-19 pandemic, with approximately half of the subjects (*n* = 28) participating in 2020 (before campus shut down), and the other half (*n* = 35) participating in 2021. Details regarding how procedures were changed to meet COVID-19 safety standards are discussed below. We report an exploratory analysis comparing data collected before and after campus shut down in the Supplemental Materials.

Recruiting subjects in the spring of 2021 proved more difficult, so we made two alterations to our pre-registered plans for recruitment and exclusions. First, we began offering cash payment as an additional option in place of course credit. The majority of subjects (*n* = 51) were compensated with course credit, and only a small subset of subjects opted for a $10 cash payment (*n* = 12). Additionally, in order to retain more subjects in the sample we decided to only remove subjects with more than 20% of trials lost due to blinking (not 20% of trials lost due to blinking and incorrect responses combined).

After replacing subjects who were excluded due to experiment or equipment malfunction (11 subjects) and blinking-related data loss (two subjects), we met our target sample size of 50 participants (38 female, 12 male; Age *M* = 19.62, *SD* = 1.09). The sample size for Experiment 2 was selected based on sufficient power to detect effects in Experiment 1. All subjects were native speakers of English with normal hearing and normal (or corrected-to-normal) vision, and had no extensive exposure to Mandarin Chinese.

Materials

Stimuli for Experiment 2 included recordings of two native, American-English speakers and two nonnative, Mandarin Chinese-accented speakers of English reading sentences with four keywords each (from the same sentence set as McLaughlin & Van Engen, 2020; Van Engen et al., 2012).[2] All of the speakers were female. None of the speakers from McLaughlin and Van Engen (2020) were included. The two nonnative speakers were selected from a set of seven speakers (all native speakers of Mandarin) after an online transcription pilot. Nonnative Speaker 1 was found to be 93% intelligible in quiet and Nonnative Speaker 2 was found to be 94% intelligible in quiet.

When examining the time-course of listening effort with pupillometry, it is important to match speaking rate across conditions (McLaughlin & Van Engen, 2020). Thus, in order to match rate across speakers in Experiment 2, the nonnative speakers were instructed to read naturally while the native speakers were instructed to read slightly slower than natural. The full set of recordings contained 220 target sentences per speaker, and when selecting target sentences for the present experiment (which required a total of 104 targets), we matched the average lengths of target files across speakers (2860 ms). We also aimed to select the sentences with the highest intelligibility across the two nonnative speakers.[3]

Procedure

Participants entered the lab and confirmed that they were native English speakers, did not have extensive exposure to Mandarin-accented speech (e.g., living with a Mandarin speaker, studying Mandarin), had normal hearing, and had normal or corrected-to-normal vision. For pre-

[2] While we chose Mandarin as the nonnative accent in the current study, it is important to note that nonnative accents are not homogenous. The effects observed in the present study using American English and Mandarin may not be representative of all accents. It remains an open question whether similar patterns will be observed for across-accent switching between different native and nonnative accents.

[3] It should be noted that we could not examine sequence effects in task performance (e.g., RT or accuracy) because the present paradigm requires high levels of intelligibility for all speakers. These kinds of sequential adjustments remain an intriguing avenue for future research.

COVID participants, the experimenter brought them to a testing room and began the instructions. For the COVID protocol, participants were instructed to enter the testing room and instructions were delivered by an ongoing video call. The trial procedure was adapted from McLaughlin and Van Engen (2020).

Participants wore circumaural headphones and rested their chins on a head-mount that was 90 cm away from a 53.5 cm by 30 cm computer screen. All equipment was positioned following EyeLink specifications. A nine-point calibration and validation procedure was conducted for all subjects before they began the task.

During the task, participants were instructed to fixate on a cross located in the center of the screen. The color of the cross was used to signal which part of the trial they were in. When the cross was red, participants were instructed to reduce blinking as much as was comfortable and to attend to the auditory stimulus. When the cross was blue, participants were instructed to blink freely. Each trial began with a baseline period of 3000 ms of silence and a red cross. Next, with the red cross still present, the stimulus played followed by a delay period of 3000 ms. At this point, the color of the cross turned to blue, indicating that subjects could blink freely. Participants were instructed to repeat what they heard aloud. For pre-COVID participants, responses were recorded with an audio recorder. For COVID participants, responses were recorded as part of the ongoing video call. Finally, participants pressed the spacebar to move to the next trial, and a 3000 ms silent delay period with a blue fixation cross was presented. This delay allowed the pupil response to recover between trials.

Participants began with four practice trials, one per speaker. These practice trials followed the same trial procedure as the experimental task. Next, subjects completed the four 25-trial experimental blocks. Each block began with a start trial, which was not included in our analyses because it was neither a repeat nor a switch trial. Each of the four speakers was the start trial in one of the four blocks. Of the remaining 24 trials per block, each of the four speakers was presented six times. The order of trials was unpredictable from the participant's perspective, but was pseudorandom so that the lists contained the same number (eight) of the key transition types: no switch, within-accent switch, and across-accent switch. For no switch transitions, the speaker from trial N-1 spoke on trial N (i.e., the current trial). For within-accent switches, the speaker on trial N-1 had the same accent as the speaker on trial N, but was a different speaker. Lastly, for across-accent switches, the speaker on trial N-1 was a different speaker with a different accent. Additionally, for each of the three main conditions, we considered the accent of the speaker on trial N. We compared, for example, an across-accent switch from native to nonnative with an across-accent switch from nonnative to native. Between each block, there was a self-timed break.

After completing the four experiment blocks, participants completed language and demographic questionnaires. Finally, participants were debriefed on the task. The entire procedure took approximately 45 minutes.

Data preparation

Repetitions of the target sentences were scored to determine whether subjects correctly understood the speaker. Each sentence had four keywords (e.g., the *gray mouse ate* the *cheese*). Any trial in which any keyword was misidentified or missing was excluded from analyses. Differences in plurality and verb tense (specifically differences in use of -ed morpheme) were allowed. In the nonnative accent condition, 7.7% of trials were removed for being <100% intelligible (or in some cases, due to poor recording quality that prevented response scoring). In the native accent condition, only 1.2% of trials were lost.

For the pupil data, pre-processing was completed using the R package gazeR (Geller, Winn, Mahr, & Mirman, 2020). Subjects with more than 20% data loss due to blinking were excluded (two subjects). Periods of missing data due to blinks were next identified and extended 100 ms prior and 200 ms following. This process removes extraneous values that occur when the eyelid is partially obscuring the pupil. For these extended blink windows, linear interpolation was used to fill in the missing data. A five-point moving average then smoothed the data. The median pupil diameter during the 500 ms immediately preceding stimulus onset was used as the baselining value for each trial. Subtractive baselining was used (Reilly et al., 2019). As a final step, data was time-binned, reducing the sampling frequency from 500 Hz to 50 Hz.

Analysis window selection

Time window selection for growth curve analysis can increase researcher degrees of freedom during the analysis process (Peelle & Van Engen, 2021). To avoid biasing our analyses, we selected our analysis window without viewing its influence on the effects of interest. The only data viewed prior to window selection was a single plotted curve summarizing the mean of all trial and subject data (i.e., to confirm a polynomial analysis was appropriate). Due to the delay of the pupil response, which is typically 200-300 ms, we opted to begin our analysis window at 300 ms after target onset. The end of the analysis window was based on the average offset time of the stimuli (2860 ms).

Results

The random effect structure matched that of Experiment 1.[4] Table 2 summarizes all log-likelihood model comparisons from the growth curve analysis of the full dataset. The linear, quadratic, and cubic polynomials all significantly improved fit (all *p*’s < .001). Given the

[4] As the fixed effects in the model grew more complex, it became more difficult for models to converge and we had to simplify the random effect structure by removing the polynomial random slopes nested with items.

complex shape of the pupil response, we also tested whether a quartic polynomial would improve fit. It did not ($\chi^2(1) = 0.02$, $p = .88$), and thus was not retained in subsequent models.

Table 2

Log-likelihood Model Comparisons for Growth Curve Analysis of Experiment 2

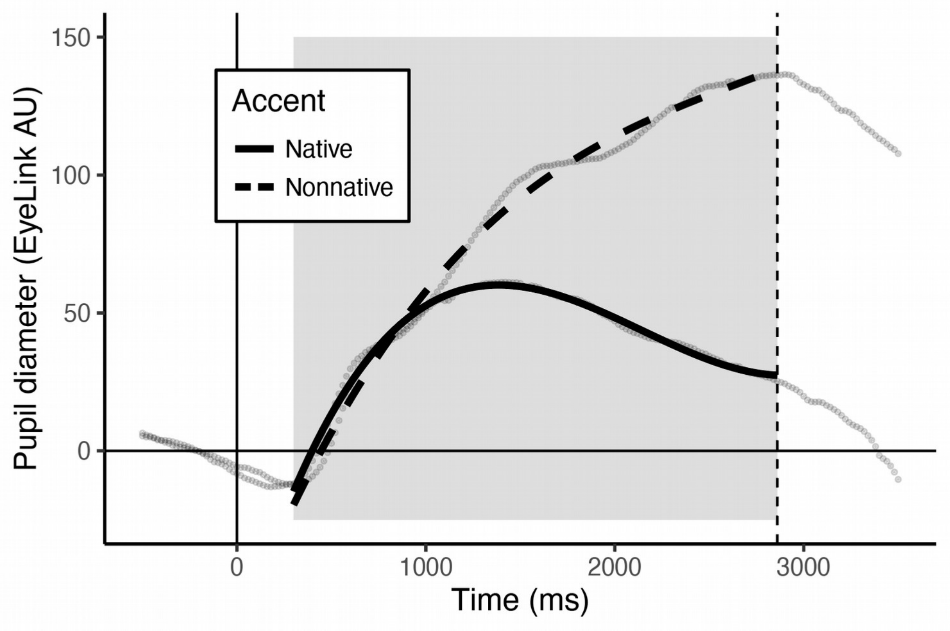
Effect	χ^2	df	*p*
Linear polynomial	3074.30	1	< .001 ***
Quadratic polynomial	1147.70	1	< .001 ***
Cubic polynomial	112.64	1	< .001 ***
Quartic polynomial*	0.02	1	.88
Accent (Levels: Native, Nonnative)	2877.50	1	< .001 ***
Switch (Levels: No Switch, Within-accent Switch, Across-accent Switch)	272.91	1	< .001 ***
Accent x Linear polynomial	2467.20	1	< .001 ***
Accent x Quadratic polynomial	37.99	1	< .001 ***
Accent x Cubic polynomial	26.41	1	< .001 ***
Switch x Linear polynomial	28.96	2	< .001 ***
Switch x Quadratic polynomial	16.40	2	< .001 ***
Switch x Cubic polynomial	3.82	2	.15
Accent x Switch	109.34	2	< .001 ***

Note. *Effect not retained in subsequent models.

First, we examined the dummy-coded main effects of accent (reference level: native) and switch (reference level: no switch). Log-likelihood model comparisons indicated that both accent ($\boldsymbol{\chi}^2(1) = 2877.5$, p < .001) and switch ($\boldsymbol{\chi}^2(2) = 272.91$, $p < .001$) significantly improved fit. Model estimates indicated that there was larger pupil response for nonnative-accented compared to native-accented speech ($\boldsymbol{\beta} = 42.17$, $p < .001$; Figure 2), and larger pupil response for switching speakers within accent ($\boldsymbol{\beta} = 8.37$, $p < .001$) and across accent ($\boldsymbol{\beta} = 16.06$, $p < .001$) as compared to listening to the same speaker consecutively (Figure 3). Interactions of accent and switch with the polynomial time terms are reported in Table 2.

Figure 2

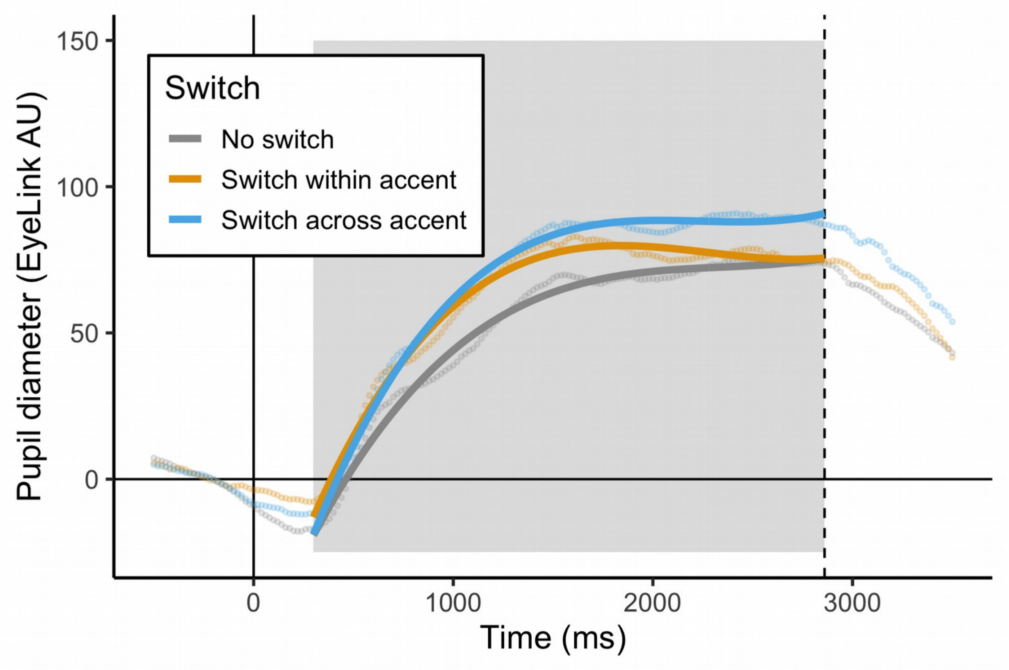
Effect of Accent in Experiment 2



Note. The effect of accent on the size of the pupil is shown with model fits and raw data points. The y-axis shows pupil diameter in EyeLink AU (Arbitrary Units), where zero is the baseline calculated to align data across trials. The x-axis shows time in ms, beginning at trial start (zero). The dashed vertical line indicates the average offset time for all stimuli. The gray box indicates the window of the data used in analyses.

Figure 3

Effect of Switch in Experiment 2

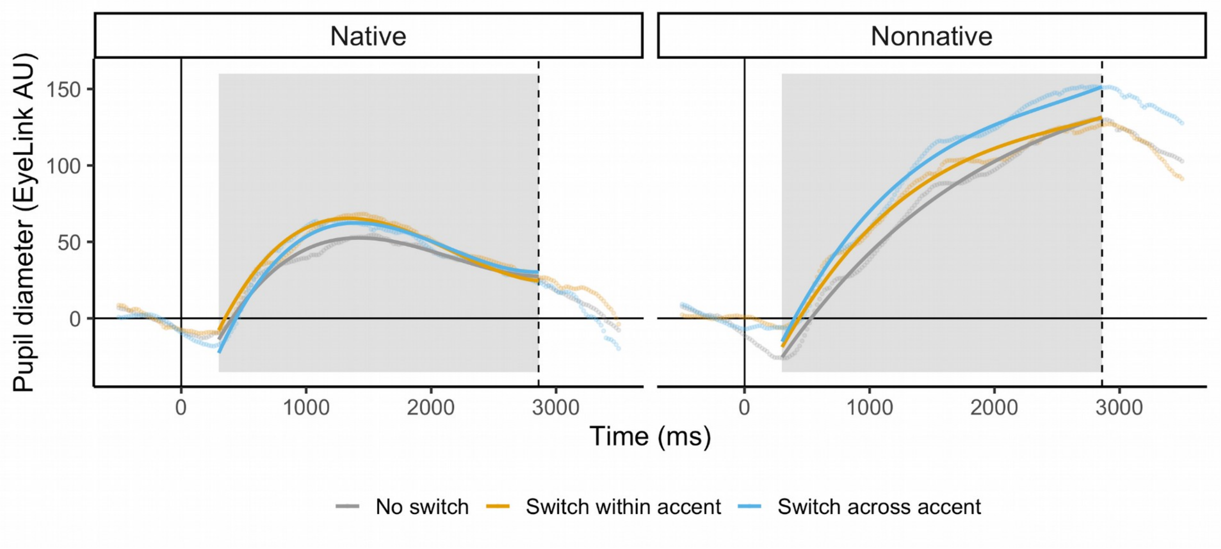


Note. The effect of switch on the size of the pupil is shown with model fits and raw data points. The y-axis shows pupil diameter in EyeLink AU (Arbitrary Units), where zero is the baseline calculated to align data across trials. The x-axis shows time in ms, beginning at trial start (zero). The dashed vertical line indicates the average offset time for all stimuli. The gray box indicates the window of the data used in analyses.

Next, we examined the interaction between accent and switch. Adding the interaction significantly improved model fit ($\chi^2(2) = 109.34$, $p < .001$), but model estimates indicated that there was no difference in the cognitive demands for the within-accent switch condition (as compared to the no switch condition) for native and nonnative accents ($\beta = -1.99$, $p = .31$). Rather, this interaction appears to be driven by across-accent switching, for which pupil response is greater when switching from native to nonnative accent ($\beta = 16.32$, $p < .001$; Figure 4).

Figure 4

Interaction of Accent and Switch Effects in Experiment 2



Note. The interaction between the effects of accent and switch is captured in two panels, with the labels indicating the accent of the current trial. For both panels, the y-axis shows pupil diameter in EyeLink AU (Arbitrary Units), where zero is the baseline calculated to align data across trials. The x-axis shows time in ms, beginning at trial start (zero). The dashed vertical line indicates the average offset time for all stimuli. The gray box indicates the window of the data used in analyses.

To directly examine the cognitive demands for within- versus across-accent switching, we reordered the levels of switch to make the within-accent switch condition the reference level of the factor. When examining a model without interactions, model estimates indicated that pupil response was greater for switching speakers across-accent than within-accent ($\boldsymbol{\beta}$ = 7.69, *p* < .001; Figure 2). Additionally, when examining a model with all of the fixed effects including interactions, model estimates from the accent by switch interaction indicated that the difference in pupil response between the across-accent and within-accent conditions was larger for the nonnative than the native condition ($\boldsymbol{\beta}$ = 18.31, p < .001; Figure 3).

We also pre-registered follow-up analyses for directly examining switching costs for the native and nonnative accent conditions separately. We began by subsetting the data so that only trials from the native accent condition remained. Log-likelihood model comparisons indicated that switch improved model fit ($\boldsymbol{\chi}^2(2) = 58.09$, p < .001). Model estimates indicated that pupil response for switching both within- ($\boldsymbol{\beta}$ = 8.70, p < .001) and across-accent ($\boldsymbol{\beta}$ = 8.65, p < .001) was greater than pupil response for not switching. We next reordered the levels of switch (reference level: within-accent switch) to directly compare the pupil response for within- and across-accent switching. The model estimate indicated that there was not a significant difference in pupil response for within- and across-accent switching for native accent ($\boldsymbol{\beta}$ = -0.04, p = .97; Figure 4).

Finally, we completed this process again for the nonnative accent data. For nonnative speech, switch also improved model fit ($\boldsymbol{\chi}^2(2) = 388.17$, p < .001), and model estimates indicated that within- and across-accent switching resulted in larger pupil response than listening to the same nonnative speaker consecutively ($\boldsymbol{\beta}$ = 5.78 and $\boldsymbol{\beta}$ = 26.45, respectively; both p's < .001). After reordering the levels of switch (reference level: within-accent switch) we were able to confirm that pupil responses were larger for across-accent switching than within-accent switching ($\boldsymbol{\beta}$ = 20.67, p < .001), as expected based on the interaction between accent and switch discussed above (Figure 4).

Discussion

The results of Experiment 2 replicate and extend the findings of Experiment 1. As predicted, trials with speaker switches were more difficult than those with speaker repetitions, across-accent switches were more difficult than within-accent switches, and an asymmetry emerged – switching from native to nonnative accent was more demanding than switching from nonnative to native accent. Lastly, Experiment 2 also replicated McLaughlin and Van Engen (2020), demonstrating that perceiving highly-intelligible nonnative accent can be more demanding than perceiving native accent.

General Discussion

Across two experiments, we presented novel evidence of sequence effects on the cognitive demands of speech perception. Using pupillometry as an online measurement of cognitive demand, we found that switching between speakers universally increased demand. Additionally, switching speakers across accents increased demands more than switching speakers within an accent – although this effect was qualified by an interaction such that switching from a native to a nonnative speaker was more costly than switching between two nonnative speakers whereas switching from a nonnative to a native speaker was not more costly than switching between two native speakers. These results reveal that listeners make dynamic and cognitively effortful adjustments in response to different speakers.

The present study does not yet identify the mechanism(s) that underlie these sequence effects for speech perception. One possibility would be a linguistic mechanism that supports speaker and accent accommodation. Classic linguistic accounts suggested that differences among talkers must be *normalized* by listeners, perhaps with a mechanism that uses a talker's vocal blueprint to calibrate internal phonetic categories (Gerstman 1968; Ladefoged and Broadbent 1957; Nusbaum and Morin 1992).[5] Talker-switching costs found in the present study and the multi-talker processing costs literature (Carter et al., 2019; Choi & Perrachione, 2019; Heald & Nusbaum, 2014; Magnuson et al., 2021; Mullennix et al., 1989; Saltzman et al., 2021; Wong et

[5] Many papers examining multi-talker processing costs discuss talker-switching costs in both the context of a normalization mechanism as well as exemplar theory, although in some cases the data has favored a normalization mechanism account (e.g., Magnuson et al., 2021).

al., 2004) indicate that alternating between speakers is a cognitively demanding process. Alone, this finding is easily explained by the engagement of a talker normalization mechanism. However, it remains unclear whether a talker normalization mechanism can also explain the asymmetry found in the present study for native-to-nonnative versus nonnative-to-native switching. Why would adjusting one's perceptual categories from a native accent to a nonnative accent be more demanding than switching between two nonnative-accented speakers, while adjusting from a nonnative to a native accent is no different than switching between two native-accented speakers? The 'phonetic distance' between the two accents is equivalent, and yet the cognitive processing costs are asymmetrical.

These switching effects between speakers and accents might be explained by dynamic, trial-by-trial adjustments triggered by differing effort needed to perceive each speaker. When a nonnative speaker occurs on trial N-1, the relatively greater upregulation of effort benefits perceiving a nonnative speaker on the next trial (trial N), and does not harm performance when switching to a native speaker. This accounts for our finding that the nonnative-to-native switches were no more costly than the native-to-native switches. In contrast, when a native speaker occurs on trial N-1, the relatively reduced upregulation of effort is presumably inadequate to accommodate switching to a nonnative speaker on the next trial, in this case resulting in a greater cost for native-to-nonnative switches than nonnative-to-nonnative switches.

A related but distinct way to conceptualize the trial-by-trial adjustments is based on differences in the upregulation of control[6] (cf. Botvinick et al., 2001). Here, one can ask whether the sequential adjustments are proactive (i.e., the accent on trial N-1 heightens control that is

[6] In the CSE literature, researchers have distinguished accounts based on bottom-up associative mechanisms (e.g., benefit is observed for consecutive incongruent trials because stimulus and/or response features repeat; e.g., Mayr et al., 2001) from control mechanisms (i.e., benefit is observed because the conflict-triggered heightening of attention toward the goal-relevant dimension on N-1 benefits trial N; e.g., Botvinick et al., 2001). Given the current study's use of unique sentences and responses on each trial, an account based on bottom-up associative mechanisms appears unlikely. Rather, a control-based account is more plausible.

actively maintained and influences trial N) or reactive (i.e., the heightening of control might residually carry over to the next trial with such carry over anticipated to decay at longer intervals; e.g., Scherbaum et al., 2012). Distinguishing these forms of adjustments is critical to understanding specific mechanisms that support speaker and accent accommodation. In the control literature, CSEs have been interpreted as reactive because they are diminished or eliminated for “long” ISIs between 2250 and 5000 ms (Egner et al., 2010, but see Duthoo et al., 2014b). The sequence effects in the present study were observed using inter-stimulus intervals that were quite long (10 - 20 seconds), thus favoring a proactive account. However, a distinct reactive account merits consideration—participants might associate the accent on trial N-1 with the degree of control used on that trial (e.g., nonnative = high control), and the associated degree of control might be reactivated when encountering the accent on trial N (i.e., a learning account; Freund & Nozari, 2018). This type of reactive mechanism appears to persist across long delays and intervening trials, consistent with the present findings (Freund & Nozari, 2018). A key avenue for future research will be determining the nature of the sequential adjustments that occur for multi-talker and multi-accent perception.

Conclusion

The challenge listeners face when mapping acoustic input onto linguistic representations can be complicated by both speaker and accent variability. In the present study, we investigated the cognitive demands associated with accommodating speaker and accent variation. To this end, we used pupillometry to track cognitive processing load for trial-to-trial switches between speakers of the same or different accents. Our results indicated a universal cost for switching between speakers, and an asymmetry in the costs for switching between different accents (i.e., native vs. nonnative accents). Specifically, switching from a native speaker to a nonnative speaker was especially cognitively demanding (as compared to switching between two nonnative speakers), while the reverse was no different than switching between two native speakers. These sequence effects observed for speech perception align with work examining multi-talker processing costs, and provide novel evidence that listeners actively engage cognitive resources to accommodate speaker and accent variability. Further, these findings align with research on sequential adjustments in cognitive control, demonstrating that prior context can affect the current demands of a cognitive task.

Declarations

Funding

This work was supported by National Science Foundation Graduate Research Fellowship DGE-1745038 (Drew J. McLaughlin).

Conflicts of interest/Competing interests

The authors have no relevant financial or non-financial conflicts of interest to disclose.

Ethics approval

Approval was obtained from the ethics committee of Washington University in St. Louis.

Consent to participate

All subjects consented to participation in the study.

Consent for publication

Not applicable.

Availability of data and materials

The datasets generated during and/or analyzed during the current study are available in the Open Science Framework repository, https://osf.io/nzgyb/files/.

Code availability

The code used to analyze data from the current study are available in the Open Science Framework repository, https://osf.io/nzgyb/files/.

Authors' contributions

Drew J. McLaughlin: conceptualization, design, implementation, data management, data analysis, writing and editing

Jackson S. Colvett: conceptualization, design, implementation, data management, writing and editing

Julie M. Bugg: conceptualization, design, writing and editing, oversight

Kristin J. Van Engen: conceptualization, design, writing and editing, oversight

Open Practices Statement: *The data and materials for all experiments are available at* https://osf.io/nzgyb/. *Only Experiment 2 was preregistered.*

References

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting Linear Mixed-Effects Models using lme4. In *arXiv [stat.CO]*. arXiv. http://arxiv.org/abs/1406.5823

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

Brown, V. A., McLaughlin, D. J., Strand, J. F., & Van Engen, K. (2020). Author accepted manuscript: Rapid adaptation to fully intelligible nonnative-accented speech reduces listening effort. *Quarterly Journal of Experimental Psychology*, 1747021820916726.

Carter, Y. D., Lim, S.-J., & Perrachione, T. K. (2019). Talker continuity facilitates speech processing independent of listeners' expectations. *19th International Congress of Phonetic Sciences*. https://sites.bu.edu/cnrlab/files/2019/05/Carter-Lim-Perrachione-2019-ICPhS-ERTA.pdf

Choi, J. Y., & Perrachione, T. K. (2019). Time and information in perceptual adaptation to speech. *Cognition*, *192*, 103982.

Duthoo, W., Abrahamse, E. L., Braem, S., Boehler, C. N., & Notebaert, W. (2014a). The congruency sequence effect 3.0: a critical test of conflict adaptation. *PloS One*, *9*(10), e110462.

Duthoo, W., Abrahamse, E. L., Braem, S., Boehler, C. N., & Notebaert, W. (2014b). The heterogeneous world of congruency sequence effects: An update. *Frontiers in Psychology*, *5*, 1001, doi:10.3389/fpsyg.2014.01001

Egner, T. (2007). Congruency sequence effects and cognitive control. *Cognitive, Affective &*

Behavioral Neuroscience, *7*(4), 380–390.

Egner, T., Ely, S., & Grinband, J. (2010). Going, going, gone: characterizing the time-course of congruency sequence effects. *Frontiers in psychology, 1*, 154.

Gratton, G., Coles, M. G., & Donchin, E. (1992). Optimizing the use of information: strategic control of activation of responses. *Journal of Experimental Psychology: General*, *121*(4), 480–506.

Heald, S. L. M., & Nusbaum, H. C. (2014). Talker variability in audio-visual speech perception. *Frontiers in Psychology*, *5*, 698.

Magnuson, J. S., Nusbaum, H. C., Akahane-Yamada, R., & Saltzman, D. (2021). Talker familiarity and the accommodation of talker variability. *Attention, Perception & Psychophysics*, *83*(4), 1842–1860.

McLaughlin, D. J., & Van Engen, K. J. (2020). Task-evoked pupil response for accurately recognized accented speech. *The Journal of the Acoustical Society of America*, *147*(2), EL151–EL156.

Mirman, D. (2016). *Growth Curve Analysis and Visualization Using R*. CRC Press.

Mullennix, J. W., Pisoni, D. B., & Martin, C. S. (1989). Some effects of talker variability on spoken word recognition. *The Journal of the Acoustical Society of America*, *85*(1), 365–378.

Peelle, J. E., & Van Engen, K. J. (2021). Time Stand Still: Effects of Temporal Window Selection on Eye Tracking Analysis. *Collabra. Psychology*, *7*. https://doi.org/10.1525/collabra.25961

Porretta, V., & Tucker, B. V. (2019). Eyes Wide Open: Pupillary Response to a Foreign Accent Varying in Intelligibility. *Frontiers in Communication*, *4*, 8.

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.

Saltzman, D., Luthra, S., Myers, E. B., & Magnuson, J. S. (2021). Attention, task demands, and multitalker processing costs in speech perception. *Journal of Experimental Psychology: Human Perception and Performance*, *47*(12), 1673–1680.

Schmidt, J. R., & Weissman, D. H. (2014). Congruency sequence effects without feature integration or contingency learning confounds. *PloS One*, *9*(7), e102337.

van der Wel, P., & van Steenbergen, H. (2018). Pupil dilation as an index of effort in cognitive control tasks: A review. *Psychonomic Bulletin & Review*, *25*(6), 2005–2015.

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.

van Steenbergen, H., & Band, G. P. H. (2013). Pupil dilation in the Simon task as a marker of conflict processing. *Frontiers in Human Neuroscience*, *7*, 215.

Weissman, D. H., Hawks, Z. W., & Egner, T. (2016). Different levels of learning interact to shape the congruency sequence effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(4), 566–583.

Weissman, D. H., Jiang, J., & Egner, T. (2014). Determinants of congruency sequence effects without learning and memory confounds. *Journal of Experimental Psychology: Human Perception and Performance*, *40*(5), 2022–2037.

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.

Wong, P. C. M., Nusbaum, H. C., & Small, S. L. (2004). Neural bases of talker normalization. *Journal of Cognitive Neuroscience*, *16*(7), 1173–1184.

Zekveld, A. A., & Kramer, S. E. (2014). Cognitive processing load across a wide range of listening conditions: insights from pupillometry. *Psychophysiology*, *51*(3), 277–284.

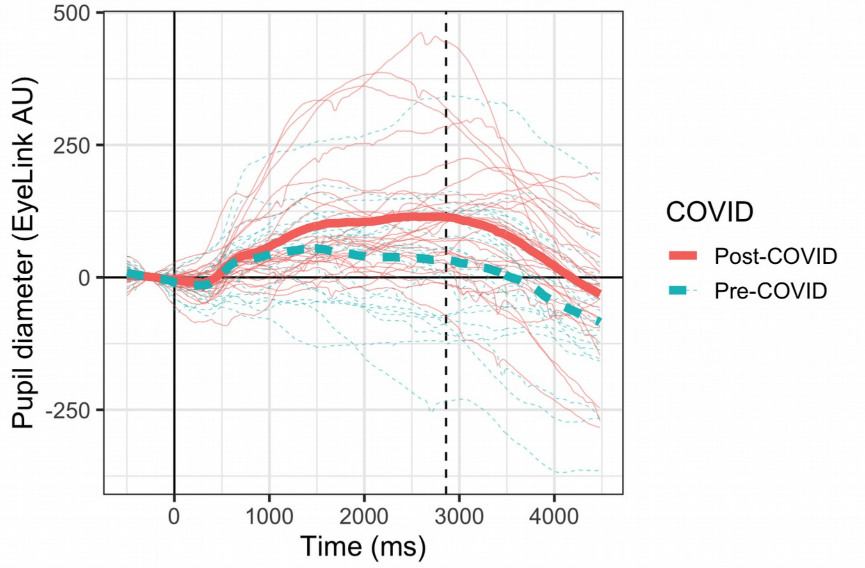
Zekveld, A. A., Kramer, S. E., & Festen, J. M. (2010). Pupil response as an indication of effortful listening: the influence of sentence intelligibility. *Ear and Hearing*, *31*(4), 480–490.

Supplemental Materials

As part of our visual inspection of the data, we compared performance of subjects who participated in the study before the onset of the pandemic ("Pre-COVID"; *n* = 22) versus subjects who participated after the onset of the pandemic ("Post-COVID"; *n* = 28). Supplemental Figure 1 shows the overall difference between these two groups and individual subject trends. A basic linear mixed-effects model with random intercepts for subjects indicated that subjects who participated after the onset of the pandemic had larger overall pupil response during the task (*p* = .04).

Supplemental Figure 1

Time of Participation Relative to the Onset of the Pandemic



Note. Overall pupil response during the task is plotted based on time of participation relative to the COVID-19 pandemic. The y-axis shows the size of the pupil over time within a trial (x-axis). Thick lines show group means, and thin lines show individual subjects' mean curves. The solid vertical line indicates the beginning of stimulus presentation and the dashed vertical line indicates the average offset of stimulus presentation.

The difference between the subjects who participated before and after the onset of the COVID-19 pandemic is surprising when considering that our equipment and experiment remained identical. One possible explanation for the data is the change to our procedures for COVID-related safety. When data collection resumed (i.e., after the suspension of in-person data collection due to COVID-19 had been lifted), we changed our procedures in the following ways: Subjects were required to wear masks at all times, including when completing the task and resting their chin on the pupillometry head mount; subjects did not interact with the researcher directly, but were communicated with via a Zoom call on a computer added to the pupillometry testing suite; and subjects were told that the Zoom call would run throughout the entire experiment, and that the researcher could hear them if they needed assistance at any point. It is reasonable to suggest that these factors could have affected engagement with the task.

Another explanation for the effect of the pandemic on the data would be differences in subjects' stress and arousal levels. Given that the pupil is affected by stress and arousal, it is plausible that subjects entered the experiment session in a heightened state of stress unrelated to the task itself (for example, worrying about coming into contact with someone sick). Indeed, for many of the "Post-COVID" subjects in the study it was likely one of their first outings since the beginning of the COVID-19 pandemic. In either case, this change could have affected pupil response during the task, resulting in larger overall pupil responses for subjects participating after the onset of the pandemic.

Although the effects of the pandemic on the current data are certainly notable, we caution against drawing conclusions from these data because of the smaller sample size in this between-subject comparison. We report these data here for the purposes of transparency as well as to alert other researchers who have used pupillometry throughout the onset and continuation of the pandemic.