

ORIGINAL ARTICLE

The role of phonology-to-orthography consistency in predicting the degree of pupil dilation induced in processing reduced and unreduced speech

Yoichi Mukai^{1,2,*} , Juhani Järvikivi² and Benjamin V. Tucker^{2,3} 

¹Modern Languages Studies Department, Vancouver Island University, Nanaimo, British Columbia, Canada,

²Department of Linguistics, University of Alberta, Edmonton, Alberta, Canada and ³Department of Communication Sciences and Disorders, Northern Arizona University, Flagstaff, Arizona, United States

*Corresponding author: Email yoichi.mukai@viu.ca

(Received 21 June 2022; revised 23 February 2023; accepted 10 April 2023; first published online 22 May 2023)

Abstract

The relationship between the ways in which words are pronounced and spelled has been shown to affect spoken word processing, and a consistent relationship between pronunciation and spelling has been reported as a possible cause of unreduced pronunciations being easier to process than reduced counterparts although reduced pronunciations occur more frequently. In the present study, we investigate the effect of pronunciation-to-spelling consistency for reduced and unreduced pronunciations in L1 and L2 listeners of a logographic language. More precisely, we compare L1 and L2 Japanese listeners to probe whether they use orthographic information differently when processing reduced and unreduced speech. Using pupillometry, the current study provides evidence that extends the hypothesis about the role of orthography in the processing of reduced speech. Orthographic realization matters in processing for L1 and L2 advanced listeners. More specifically, how consistent the orthographic realization is with its phonological form (phonology-to-orthography consistency) modulates the extent to which reduced pronunciation induces additional processing costs. The results are further discussed in terms of their implications for how listeners process reduced speech and the role of the orthographic form in speech processing.

Keywords: spoken word recognition; pronunciation-to-spelling consistency; spontaneous speech; pupillometry; generalized additive modelling

A growing body of literature has investigated the effect of spelling on processing of first (L1) and second language (L2) speech (e.g., Frauenfelder et al., 1990; Morais et al., 1979; Seidenberg & Tanenhaus, 1979; Zou et al., 2012) (See Hayes-Harb and Barrios (2021) for an overview of L2 studies). The processing of spoken words

© The Author(s), 2023. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

is affected by inconsistencies between pronunciation and spelling (e.g., Ziegler et al., 2004). Work by Ziegler et al. (2008) found that auditory lexical decisions are faster and more accurate for consistent words whose rhyme can be spelled in only one way (e.g., /-ʌk/ in *luck*) than for inconsistent words whose rhyme can be spelled in multiple ways (e.g., /-ip/ in *leap* or *keep*). A consistent relationship between pronunciation and spelling has been reported as a possible cause of unreduced pronunciations being easier to process than reduced counterparts, despite reduced pronunciations occurring more frequently (Ranbom & Connine, 2007). The present study examines the extent to which orthography affects the processing of reduced pronunciations. More specifically, how pronunciation-to-spelling consistency interacts with reduced pronunciations for L1 and L2 listeners of Japanese, a non-alphabetic language.

Orthographic consistency effect in spoken word recognition

Considerable evidence for orthographic effects in spoken word processing has been provided by studies on pronunciation and spelling consistency. In particular, pronunciation-to-spelling (i.e., phonology-to-orthography) consistency has been shown to affect the ease of spoken word processing. Research has shown slower response times and higher error rates for inconsistent words than consistent words in French (Ziegler & Ferrand, 1998) and English (Ziegler et al., 2008) in auditory lexical decision. Furthermore, Ziegler et al. (2004) revealed an effect of spelling probability within a consistency effect. They found the slowest response times and the highest error rates for inconsistent non-dominant spelling words, followed by the inconsistent dominant spelling words, and the fastest responses and lowest error rates for consistent words. They also compared the relative effect of consistency across three tasks: auditory lexical decision, rhyme detection, and auditory naming, and found that the effect is strongest in the auditory lexical decision task, the second strongest in the rhyme detection task, and the smallest (or null) in the auditory naming task. Similar results were also reported for Portuguese (Ventura et al., 2004) and Thai (Pattamadilok et al., 2008).

While it has been argued that the emergence of consistency effects is dependent on a meta-phonological or meta-linguistic analysis required to complete a given task, the effect has been demonstrated in a semantic and non-(meta)linguistic task as well. Peereman et al. (2009) employed a semantic and gender categorization task and measured response latencies and accuracy. Their results indicated that there is a consistency effect for both tasks. Similarly, Pattamadilok et al. (2009) examined semantic categorization during event-related potentials (ERPs) and a Go-NoGo task. The ERPs revealed the time-locked consistency effects, where early inconsistent words show an early effect and late inconsistent words display a late effect of inconsistency. Additionally, Perre et al. (2011) demonstrated the consistency effect during ERPs using a Go-NoGo non-(meta)linguistic task, where participants pressed a button for white noise and they did not press the button for French words. Although participants did not have to make a linguistic decision on the French words, the consistency effect emerged.

Work by Perre et al. (2009), also employed ERPs, but they used standardized low-resolution electromagnetic tomography (sLORETA) to determine the possible underlying cortical generators of the consistency effect. Using sLORETA, they directly examined the mechanism by which orthography affects spoken word processing. Their predictions were based on two accounts hypothesizing the nature of the consistency effect: (1) orthographic information is co-activated with phonological information and (2) orthographic information “contaminates” phonological representations during development. The first account predicts that orthographic information is co-activated with phonological information online. Autonomous orthographic and phonological representations are bidirectionally linked, and these two involuntarily interact with each other (Dijkstra et al., 1998; Stone et al., 1997). This *online view* is in line with the feedback loop in interactive-activation models in the resonance theory framework (McClelland & Rumelhart, 1981; Stone & Van Orden, 1994), where activation of one representation “feeds forward” to another representation and then the activated representation “feeds back” to the original representation to alter the activation level. According to this view, learning to read/write builds strong and permanent associations between phonological and orthographic representations. Accordingly, Perre et al., (2009) postulated activation in brain regions that are responsible for both phonological and orthographic information processing (See Cohen et al. (2004); Dehaene et al. (2001); Perre et al., (2009) for detailed discussion of cortical systems for the processing of phonological and orthographic information). The second account predicts that orthographic information “contaminates” phonology during the development of reading/writing skills, thus alternating the nature of the phonological representations (Muneaux & Ziegler, 2004; Ziegler & Goswami, 2005). According to this *restructuring view*, orthographic information is utilized to restructure phonological representations of lexical representations. That is, orthographically consistent words build better specified and finer-grained phonological representations than inconsistent words do in the course of learning to read and write. Accordingly, in contrast to the online view, an orthographic consistency effect emerges because of the qualitative difference between the phonological representations of consistent and inconsistent words. This account follows the conceptualization of the phonological restructuring model (Metsala, 1997; Metsala & Walley, 1998), where children recognize words in a more holistic way than adults do, but as children’s vocabulary size grows, restructuring of lexical representations progresses, resulting in lexical representations becoming more segmental for successful recognition of spoken words.

Since these two accounts are not mutually exclusive, it is possible to find evidence supporting both accounts simultaneously, which would become the third account. For this account, Perre et al., (2009) postulated activation in brain regions that are responsible for both phonological and orthographic information processing, as well as differences in activation of brain regions that are responsible for phonological processing between consistent and inconsistent words simultaneously. While Perre et al., (2009), as well as Pattamadilok et al. (2010), found evidence for the phonological restructuring account, Chen et al. (2016) also employed sLORETA and found evidence supporting the third account in Mandarin, possibly suggesting script-specific (alphabetic vs. logographic) processing for orthographic consistency effects.

Orthographic consistency effect and logographic scripts

Languages that utilize logographic scripts, such as Mandarin and Japanese, also show the consistency effect. In such languages however, the pronunciation and spelling mappings are not straightforward because, in the Japanese writing system for example, logographic characters do not reliably correspond to sub-lexical phonological units (Fushimi et al., 1999; Wydell, 1998; Wydell et al., 1993). However, logographs can be decomposed into smaller sub-characters, one of which is the phonetic radical that often provides a clue to the pronunciation of the whole character. The phonetic radicals have been utilized to define the pronunciation and orthographic consistency for Mandarin. Using ERPs and sLORETA, Chen et al. (2016) investigated the effect of orthography by measuring homophone density and orthographic consistency. The ERP measures revealed both effects of the orthographic consistency and homophone density, and the sLORETA indicated activation of both in the phonological and the orthographic processing regions. Similar results were found in Chao et al. (2021) where they employed event-related functional magnetic resonance imaging and functional connectivity analysis.

While a number of studies investigated the effect of orthography-to-phonology (O-P) consistency in Japanese (Fushimi et al., 1999; Hino et al., 2011; Saito et al., 1998; Wydell et al., 1995; Wydell et al., 1993), research on the effect of phonology-to-orthography (P-O) consistency is limited. One of the experiments in Hino et al. (2017) investigated the P-O consistency effect in Japanese using the P-O consistency index. Hino et al. (2011) calculated the index based on the frequency of phonological and orthographic neighbors of target words, which is in line with the approach taken by Fushimi et al. (1999) that follows the procedures employed in Jared et al. (1990). Hino et al. (2017) selected 48 Japanese logographic words based on the index, 24 high- and 24 low-consistency words, and conducted an auditory lexical decision task. They found that response latencies for low-consistency words are longer than for high-consistency words.

Orthographic consistency effect and reduced pronunciation

Mitterer and Reinisch (2015) extended the investigation of the orthographic effect to reduced pronunciations by looking at conversational spontaneous speech. They compared the processing cost of deletion of the orthographically uncoded segment /ʔ/ and orthographically coded segment /h/ in German and Maltese. Their hypothesis was that, if orthographic coding plays a role in the processing of spoken words, the deletion of orthographically uncoded segments should have a smaller impact on processing than the deletion of orthographically coded segments. They conducted a visual world eye-tracking experiment and pronunciation judgment task. Stimuli for the eye-tracking experiment were recorded as casually pronounced words and presented to participants with a carrier sentence containing discourse markers (e.g., *like*) and contractions (*you've*). The results indicated that the processing cost caused by the reduction of orthographically coded and uncoded segments does not differ, suggesting that orthographic coding does not affect processing. For the pronunciation judgment task, stimuli were recorded as carefully pronounced words and presented to

participants with a short carrier phrase without discourse markers or contractions. In contrast to the findings in the eye-tracking experiment, the researchers found the effect of orthographic coding of segments in reduction.

As a result, Mitterer and Reinisch (2015) concluded that orthography does not play an important role in processing conversational-style speech. This result is in line with research arguing that careful speech can induce participants to use orthographic information to perform tasks, enhancing task effects (Bates & Liu, 1996; Cutler et al., 2010; McQueen, 1996; Titone, 1996). Mitterer and Reinisch (2015) also noted a few issues to be addressed in future work. One of the relevant issues is to further describe how task type and speech style interact to activate orthography, as the contrast between their two experiments confounds speech styles (casual vs. careful) with task types (explicit vs. implicit): They found an effect of orthography with the pronunciation judgment task (explicit task) using careful speech, but not with the visual world paradigm (implicit task) using casual speech.

In casual and spontaneous speech, reduction is ubiquitous and phonetic realization varies substantially. Segments and/or syllables can be realized with varying degrees of shortening, deletion, and/or incomplete articulation compared to a dictionary entry for the word's pronunciation (Ernestus & Warner, 2011; Greenberg, 1999; Warner & Tucker, 2011). For example, in English, *yesterday* may be realized as [jɛʃeɪ] (Tucker, 2007) and in Japanese, *sukoshi hanashite* "speak a little bit" as [sukofanaʃite] (Arai, 1999).

The important question for our research is how such variant pronunciations are realized. Although these reduced pronunciations occur more frequently than unreduced dictionary entry pronunciations, a number of studies found that these reduced forms are recognized less efficiently than their unreduced counterparts (e.g., Arai et al., 2007; Brouwer et al., 2013; Connine & Pinnow, 2006; Ernestus et al., 2002; Janse et al., 2007; van de Ven et al., 2011). For example, Tucker (2011) found that the comprehension of words with reduced word-medial [r] and /g/ had slower response latencies than words with unreduced counterparts. Similar results were found in Ranbom and Connine (2007), where the more frequent pronunciation of reduced variant [r] elicits slower response latencies than the less frequent pronunciation of careful spelling-like variant /nt/. While Pitt (2009) argues that the reason that /nt/ is easier to recognize than [r] may be because /nt/ is perceptually more distinct than [r], Ranbom and Connine (2007) and others argue that the reason that unreduced forms are easier to process than reduced counterparts could be due to the orthographic form of the words, specifically the consistent relationship between the unreduced pronunciation and its orthographic form (Charoy & Samuel, 2019; Racine et al., 2014; Viebahn et al., 2018). Crucially, these spontaneous speech studies suggest that the surface form (i.e., phonetic realization) of words could be substantially inconsistent with its spelling, meaning that the inconsistency between the way *yesterday* is pronounced and spelled should be greater for the reduced pronunciation [jɛʃeɪ] than for the unreduced one /jɛstɚdeɪ/, and the inconsistency could affect perception of spoken words.

The present study

Previous research has demonstrated effects of orthography in spoken word processing in both alphabetic and logographic languages in both L1 and L2 listeners. In the present study, we extend this research by investigating the effect of P-O consistency for reduced and unreduced pronunciations in L1 and L2 listeners of a logographic language. More precisely, we compare L1 and L2 Japanese listeners to probe whether they use orthographic information differently when processing reduced and unreduced pronunciations. For L1 listeners, phonological representations are acquired prior to orthography, but L2 learners are likely to acquire phonological and orthographic information simultaneously or the acquisition of orthographic representation may even precede that of phonological representations. A number of studies have shown that orthographic effects appear in L2 word recognition with logographic languages (Mitsugi, 2018; Qu et al., 2018) and knowledge of spelling-pronunciation correspondences in L1 impacts the perception and learning of L2 sounds (Bassetti, 2006; Escudero et al., 2008; Escudero & Wanrooij, 2010; Han et al., 2020; Hayes-Harb et al., 2018; Hayes-Harb et al., 2010; Showalter & Hayes-Harb, 2015). The research focusing on these topics along with the methodology discussed below is unique to the present study as far as we know.

The present study employs pupillometry, the measurement of pupil dilation (Kahneman & Beatty, 1966), to investigate P-O consistency effects between reduced and unreduced pronunciations. This methodology has been applied to a variety of psycholinguistic studies (e.g., Geller et al., 2016; Guasch et al., 2017; Haro et al., 2017; Kuchinke et al., 2007; Lóo et al., 2016) as it offers a reliable measure of cognitive effort, attention, and affect imposed by different variables in speech processing (Laeng et al., 2012). The activities within the locus coeruleus are correlated with two modes of pupillary responses: tonic and phasic (Aston-Jones et al., 1999). While the tonic (or baseline) response is slow-changing and related to the state of arousal or vigilance, the phasic pupillary response is time-locked to task-related events and stimuli (Laeng et al., 2012; Papesh & Goldinger, 2015). The phasic response has been employed in a variety of speech perception and processing studies coupled with various types of tasks, with greater pupil dilation indexing greater cognitive effort: delayed naming (Goldinger et al., 1997; Papesh & Goldinger, 2012; Porretta & Tucker, 2019), speech reception threshold test (Kramer et al., 2013; Zekveld et al., 2014; Zekveld et al., 2010), spoken language comprehension (Hubert Lyall & Järvikivi, 2021), and listening to words or syllables (Hochmann & Papeo, 2014; Tamási et al., 2017) (See Schmidtke (2018) for an overview of pupillometry studies in a wide range of linguistic research).

For our experiments, pupillometry is particularly beneficial because (1) pupillary response reflects the extent to which our variables of interest impact the amount of cognitive effort in the absence of voluntary processes (Papesh & Goldinger, 2015), (2) pupillary response is physiological and automatic; therefore, cognitive effort estimated by pupil dilation can minimize the influences from task-specific strategies (Goldinger & Papesh, 2012), and (3) pupillary dilation has been reported to reveal differences in cognitive effort even when behavioral measures exhibit equivalent performance (Karatekin et al., 2004).

Following previous studies (Papesh & Goldinger, 2012; Pattamadilok et al., 2009), we use Go-NoGo and delayed naming tasks. These tasks allow us to design our experiments in such a way that (1) target trials are free of artifacts resulting from motor movements invoked by responses that possibly influence pupil dilation (Haro et al., 2017), (2) the Go-NoGo task does not require the extraction of word meaning or the explicit analysis of phonological structure or lexical status, meaning that participants do not need to make linguistically derived decisions on the target items (Pattamadilok et al., 2009), and (3) the delayed naming task allows us to examine decision-free P-O consistency effects, as naming occurs after lexical access is complete (Papesh & Goldinger, 2012). If P-O consistency effects occur solely during the perception and/or decision stages, the effects should disappear during the speech planning stage; however, if postlexical access processes are P-O consistency-sensitive, the differences should persist, meaning that P-O consistency modulates the cognitive effort required for speech planning. Additionally, since the delayed naming task requires more active involvement (listen and repeat target words) than the Go-NoGo task (passively listen to target words), we expect overall pupillary responses to be greater for the delayed naming task than for the Go-NoGo task.

If the orthographic effect does not play a role in the processing of reduced speech, as Mitterer and Reinisch (2015) propose, we should observe the P-O consistency effect in unreduced forms but not in reduced forms, meaning that inconsistent words will induce greater pupil dilation than consistent ones for unreduced words, but there will be no difference in pupil dilation between the two types of words for reduced words. Alternatively, there may be a consistency effect for both types of words, but it might influence the two types differently (e.g., the difference could be manifested by the degree and/or time course of the effect indicated by the degree and the time course of pupil dilation), suggesting that the phonetic realization of words matters for the effect (i.e., it could mean either reduction creates an additional sound-to-spelling mismatch or the consistency effect does not affect reduced forms as much as it does unreduced forms). Furthermore, the consistency effect could influence L1 and L2 listeners differently. If L2 learners have a strong orthographic component in their lexical knowledge due to the early learning of orthography, as the orthographic bias hypothesis states in Veivo et al. (2016), L2 learners might show a stronger P-O consistency effect, meaning that the differences in the degree of pupil dilation between consistent and inconsistent words might be greater for L2 listeners than L1 listeners. Moreover, the consistency effect could be modulated by their proficiency. If phonological and orthographic representations in L2 learners, especially for low proficiency learners, do not interact efficiently because these representations are not yet fully established, as suggested by Veivo and Järvikivi (2013), we should only observe a consistency effect with high proficiency learners, meaning that inconsistent words will induce greater pupil dilation than consistent ones for advanced learners, but for beginners, there will be small or no difference between inconsistent and consistent words in pupil dilation.

Method

Participants

We recruited sixty-four participants at Nagoya University in Japan, thirty-eight of whom were native speakers of Japanese (female, $n = 16$) ranging in age from 18 to 25 years old ($M = 19.7$, $SD = 1.69$) and twenty-six of whom were native speakers of English (female, $n = 16$) who speak Japanese as a second language¹, ranging in age from 19 to 34 years old ($M = 23.2$, $SD = 4.61$). All participants reported normal or corrected-to-normal vision and hearing, and performed both the Go-NoGo and delayed naming task. For each speaker group, the task presentation order was split so that half of them performed the Go-NoGo task first and the other half did the delayed naming task first. Four L1 participants and three L2 participants were excluded from our data in both tasks due to their excessive blinks and artifacts.²

Materials

We chose 226 four-mora and two-logograph words. We used the Balanced Corpus of Contemporary Written Japanese (BCCWJ) (Maekawa et al., 2014) to collect lexical information and calculate the P-O consistency index for each target word according to the criteria in Hino et al. (2017, 2011). We selected BCCWJ because it is one of the largest Japanese datasets. Also, it is composed of more recent data and has a wider range of genres than the database used in Hino et al. (2017, 2011). All of the target words contain a word-medial nasal or voiced stop (with a few words containing both of the consonants) because previous research demonstrated that both types of consonants show various forms of reduction in Japanese (Vance, 2008). For example, in the case of a word-medial voiced stop, articulation of the word-medial voiced stop in /daigaku/ “university” is approximated due to the lack of full oral closure and realized as [daiyaku], and in the extreme case, the consonant is deleted and the realization of the word becomes [daiaku] (Arai, 1999).

Our stimuli were recorded in both reduced and unreduced forms by a female native Japanese speaker. We instructed the speaker to produce the words clearly (careful speech) for unreduced forms and casually (spontaneous speech) for reduced forms. The speaker produced multiple tokens of both forms, and we selected the most natural sounding tokens. We then normalized the amplitude of the words for presentation purposes. Table 1 illustrates the acoustic properties of both reduced and unreduced forms. The intensity difference was defined as the difference between the minimum intensity of the target segment to the averaged maximum intensity of surrounding segments (Mukai & Tucker, 2017; Warner & Tucker, 2011). Overall, reduced forms have shorter duration, faster articulation rate, lower mean pitch, and smaller intensity difference. Importantly, the difference between the two forms in our study is represented in such a way that a target segment is produced as reduced or unreduced, as discussed in Tucker (2011) and Mukai (2020), rather than the absence or presence of the segment. Previous studies on the effect of orthography in processing of reduced speech treated reduction as phonological phenomena (Mitterer & Reinisch, 2015), such as the absence or presence of a word-medial schwa in French (Bürki et al., 2012; Bürki et al., 2018). While this is true for some cases,

Table 1. Mean acoustic values of stimuli in reduced and unreduced forms

Segment type	Reduction	Word Dur.	Segment Dur.	Arctic. Rate	Mean word pitch	Int. Difference
Nasal	Unreduced	0.617	0.123	4.299	223.750	9.141
Nasal	Reduced	0.451	0.093	5.842	203.005	7.906
-	Difference	$p < 0.001$	$p < 0.001$	$p < 0.001$	$p < 0.001$	$p = 0.145$
VoicedStop	Unreduced	0.606	0.049	5.083	222.072	15.024
VoicedStop	Reduced	0.459	0.039	6.725	204.335	13.403
-	Difference	$p < 0.001$	$p < 0.001$	$p < 0.001$	$p < 0.001$	$p = 0.178$

Note. Both word and segment durations are represented in seconds. Articulation rate was measured by the number of vowels per second. *P*-values indicate the probability that the difference between the two forms was significantly different from 0.

many reduction processes appear to be gradient (Bürki et al., 2011; Ernestus & Warner, 2011; Warner & Tucker, 2011).

We created four lists for each task and each list contained 150 items (5 practice words, 113 target words (reduced or unreduced forms), and 32 non-target items (i.e., pure tones for the Go-NoGo task and filler Japanese words for the delayed naming task, which were recorded together with target words). As in Perre et al. (2011), we employed a 500-ms-long pure tone as non-target items for the Go-NoGo task and set the ratio between the target and non-target trials as 70% and 30%. The same 500-ms-long pure tone was used in the delayed naming task as well. The target words were counterbalanced across both pronunciation and task, so that none of the participants heard the same word twice within the entire experiment.

Apparatus and procedure

We designed and controlled the experiment using SR Research Experiment Builder software. Participants' right eye was tracked using an EyeLink II head-mounted eye tracker (SR Research, Canada) in the pupil-only mode with a sampling rate of 250 Hz. Prior to the beginning of each experiment, we calibrated the system using a 9-point calibration procedure. We utilized Etymotic Research insert ER1 earphones to present auditory stimuli and a 1024 × 768 resolution computer screen to present a fixation cross. Participants sat on a chair in a quiet room at a distance of approximately 60 cm from the computer screen. Luminance of the room was kept constant throughout the experiment. For the delayed naming task, we used a head-mounted Countryman E6 microphone and Korg portable digital recorder to record participants' naming responses.

In the Go-NoGo task, participants looked at a fixation cross presented at the center of the screen on a gray background for 1500 ms and heard either a pure tone or Japanese word as they continued looking at the fixation cross. They responded to the pure tone by pressing a button on a Microsoft Side Winder game-pad or they did not respond to the Japanese word (passively listening). The fixation cross disappeared after the button press triggered by the pure tone or 2000 ms after the onset

of the Japanese word. To allow pupil dilation to return to baseline, a blank screen on a gray background remained for 4000 ms after the disappearance of the fixation cross (Papesh & Goldinger, 2012).

In the delayed naming task, participants looked at a fixation cross presented at the center of the screen on a gray background for 1500 ms and heard a Japanese word as they continued looking at the fixation cross. They then waited 1000 ms and heard a 500 ms-long pure tone. They then repeated what they had heard. The fixation cross disappeared 2000 ms after the onset of the Japanese word. A blank screen on a gray background remained for 4000 ms after the disappearance of the fixation cross for the pupil to settle back to baseline. In each session of both tasks, the practice items were provided at the beginning of sessions to familiarize the participants with the task. We calibrated the eye tracker before each session and after participants took a brief break (every 29 trials). We also ran drift correction at the onset of every trial. The target and non-target items were randomly assigned to each trial by the software. Each task lasted approximately 45 minutes. Before the experiment, L2-speaking participants answered a Language Experience and Proficiency Questionnaire (LEAP-Q) (Marian et al., 2007) to provide their demographic information and Japanese language experience and proficiency. LEAP-Q has been utilized in a wide range of psycholinguistic research, and it has been shown that self-reported proficiency in LEAP-Q correlates with behavioral performance data, particularly for L2 speakers, suggesting that self-reported proficiency reflects approximations of overall language skill (see Kaushanskaya et al. (2019) for an overview of LEAP-Q).

Preprocessing pupil size data

We performed data preprocessing in the statistical environment R, version 3.4.4 (R Development Core Team, 2018). After visually inspecting the range of pre- and post-marked blinks, we cleaned the data by removing 50 samples (100 ms time window) before and after the blinks. We then linearly interpolated the removed data points for each trial. When initial and/or final samples in a trial were eye blinks or their artifacts, these samples were replaced with the nearest value to complete the interpolation. We then downsampled the interpolated data to 50Hz and smoothed it using a five-point weighted moving-average smoothing function. The same interpolation and smoothing procedures were also applied to the gaze location data to use the data as a control variable. Relevant pupillary variables were computed on a trial-by-trial basis in the time window from the onset of stimulus to 2000 ms after onset for the Go-NoGo task and from the onset of stimuli to 2500 ms after onset for the delayed naming task.

Data were cleaned and checked visually on a trial-by-trial basis to detect unexpected deviations (Winn et al., 2018). The trials that contained excessive blinks and their artifacts (more than 30% of the trial) were excluded. Based on the first author's visual inspection of trials, additional trials were excluded (1) when the peak latency was shorter than 400 ms and (2) when the peak dilation was smaller than 0 or bigger than 400 in order to remove large outliers. In total, we excluded 19.7% of the data in the Go-NoGo task and 20.7% of data in the delayed naming task for L1 participants

and 23% of data in the Go-NoGo task and 13.7% of data in the delayed naming task for L2 participants.³

Statistical considerations

We applied Generalized Additive Mixed Modeling (GAMM) (Hastie & Tibshirani, 1990) to our pupil dilation data for three reasons. First, GAMM allows us to model non-linear relationships, as well as linear relationships, between dependent and independent variables (Sóskuthy, 2017; van Rij et al., 2019; van Rij et al., 2020; Wieling et al., 2016). This was important as we expected pupil size to fluctuate over time (van Rij et al., 2019). Second, GAMM can model two or more dimensional non-linear interactions of continuous variables. Third, GAMM allowed us to control for serial dependency in time series data, namely, autocorrelation (see Baayen et al. (2017) and Wood (2017) for an overview of autocorrelation in GAMM). Because of this functionality, GAMM has been utilized not only to model pupillometric data (Lõo et al., 2016; Mukai et al., 2018; Porretta & Tucker, 2019; van Rij et al., 2019), but a variety of non-linear time series data, such as electromagnetic articulography data, the position of tongue and lips during speech (Wieling et al., 2016), formant trajectory data, the time course of formant frequencies in speech (Sóskuthy, 2017), visual world eye-tracking data (Porretta et al., 2016; Veivo et al., 2016), and event-related potential data (Kryuchkova et al., 2012; Meulman et al., 2015; Porretta et al., 2017). We performed model fitting and comparisons in the statistical environment R, version 3.4.4 (R Development Core Team, 2018) using the package *mgcv* (Wood, 2017), version 1.8-23 and *itsadug* (van Rij et al., 2017), version 2.3. We followed the procedure of fitting and evaluating models illustrated in Sóskuthy (2017); van Rij et al. (2019); Wieling (2018); Wieling et al., (2016). We employed a backward selection procedure for fixed effects and a forward fitting procedure for random effects to fit the optimal model (Matuschek et al., 2017). We evaluated the contribution of input variables by χ^2 test of fREML scores using the *compareML* function. We compared the fREML score of the full model to the score of the model without one of the input variables and kept the input variables that were justified by the comparison ($p < .05$). Inclusion of interactions was also assessed by the fREML score comparison.

Variables of interest

Our dependent variable was Baseline Normalized Pupil Dilation (in the standard arbitrary unit delivered by the eye-tracking system). We calculated the baseline pupil size for each trial by averaging the pupil size in the time window from 200 ms preceding the onset of stimulus to the onset of stimulus and performed standard baseline subtraction for each trial to quantify the degree of pupil dilation. We employed a subtractive baseline correction (absolute difference) rather than divisive baseline correction (proportional difference) because percentage measures are inflated when baseline pupil size is small (Beatty & Lucero-Wagoner, 2000; Mathôt et al., 2018; Reilly et al., 2019). Our independent variables were Language, P-O Consistency Index, Reduction, and Time.

Language is a binary category, and it represents whether the participant uses Japanese as an L1 or L2. The P-O consistency index was calculated according to the criteria in Hino et al. (2017, 2011). The index ranges from 0 to 1, with 0 indicating low consistency and 1 indicating high consistency. Reduction is a binary category, and it represents whether the participant heard a reduced or unreduced stimulus. To investigate the interaction between Language and Reduction, a new factor with four levels was created and named “LanguageReduction” (L1.Unreduced, L2.Unreduced, L1.Reduced, and L2.Reduced) (see van Rij et al. (2019); Wieling (2018); Wieling et al. (2016), for an overview of how to analyze interactions in GAMM). We were also interested in the time course of processing; therefore, Time (in milliseconds) was included as a covariate.

Additionally, we included control variables: Baseline Pupil Size (same unit as Pupil Dilation), Pupil Gaze Coordinate X and Y (x- and y-axis eye gaze position on the screen in pixels), Trial Index (from 6 to 150), Word Duration (in milliseconds), Target Segment (whether the stimulus contains a word-medial voiced stop or nasal), Target Word Frequency (log-transformed), Number of Phonological Neighbors of Target Word (log-transformed), and Number of Homophones of Target Word (Z-transformed). For the Go-NoGo task, Participant’s Mean Reaction Time (in milliseconds) was also included.

The baseline pupil size was included to account for the extent to which the pupil can dilate (i.e., a larger baseline size limits the extent of dilation). Pupil Gaze Coordinates account for the possible change in pupil size caused by different gaze locations on the screen (Wang, 2011). Trial index was included to control for an effect of trial order and the fatigue of the pupil responses (McLaughlin et al., 2022). Target Segment, Word Duration, Logged Target Word Frequency, Logged Number of Phonological Neighbors of Target Word, and Standardized Number of Homophones of Target Word were included to account for the differences in the lexical items (Kuchinke et al., 2007; Porretta & Tucker, 2019). The correlation between Logged Target Word Frequency and P-O Consistency Index was weak ($r = 0.27$) but Logged Number of Phonological Neighbors was negatively correlated with P-O Consistency Index ($r = -0.66$). That is, high consistency words tend to have a small number of phonological neighbors. Therefore, P-O Consistency Index and Logged Number of Phonological Neighbors were not included in the same model. The mean reaction time for non-target items in each participant (i.e., a button press for a pure tone) was included to account for the degree to which participants were attentive to the stimuli, as the fast (more attentive) and slow reaction time (less attentive) participants could show different patterns of effects of independent variables. That is, less attentive participants may not make any effort, which could be indicated by overall weak pupillary responses, thereby diminishing the effects.

Data, stimulus lists, analysis codes, and supplementary materials of the experiments are available on the Open Science Framework at <https://osf.io/je3t4/>.

Results

In the sections that follow, we first present the statistical analysis and result of each task for both speaker groups together, with a focus on a difference between L1 and

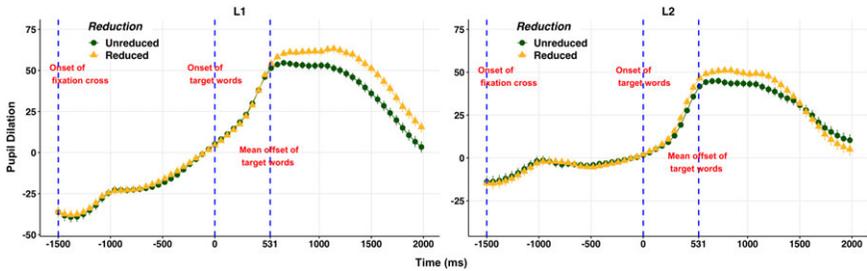


Figure 1. The grand average of pupillary dilation (y-axis) over time (x-axis) for reduced and unreduced word forms in the Go-NoGo task for L1 (left panel) and L2 (right panel) listeners. The vertical dotted line at -1500 ms indicates the onset of the fixation cross, the line at 0 ms indicates the onset of stimuli and the line at 531 ms indicates the mean offset of stimuli.

L2 listeners. We then discuss the analysis and result of each task for each speaker group separately.

Go-NoGo task

We inspected the aggregated raw pupillary dilation data prior to fitting models. Figure 1 illustrates the grand average of pupil dilation (y-axis) over time (x-axis) for reduced and unreduced word forms from -1500 ms to 2000 ms in the Go-NoGo task for L1 (left panel) and L2 (right panel) listeners. For L1 listeners, the trend of pupil dilation over time appears to be comparable between the two forms, but the reduced form demonstrates greater peak dilation and slower peak latency. For L2 listeners, both forms appear to demonstrate a similar trend to that of L1 listeners, but the difference between the two forms appears to be smaller. In addition, the overall dilation appears to be slightly greater for L1 than for L2 listeners.⁴

We chose the time window from 200 ms to 2000 ms post-stimulus onset for our analyses, as reliable effects emerge slowly in pupillary response 200 to 300 ms after a relevant cognitive event (Beatty, 1982). Pupil Dilation (dependent variable) was fitted as a function of P-O Consistency Index, LanguageReduction, and Time (independent variables). A three-way interaction was included between Time, P-O Consistency Index, and LanguageReduction. We used a tensor product smooth interaction (*te*-constructor) in order to model multi-dimensional non-linear interaction effects, allowing us to examine an effect of LanguageReduction, P-O Consistency Index and Time, as well as effects of interaction among those variables (See Sóskuthy (2017); van Rij et al. (2019, 2020); Wieling et al. (2016) for an overview of a smooth function and tensor product smooth interaction). Also, Pupil Gaze Coordinate X and Y (control variables) were included as a two-way interaction to capture the possible change in pupil size caused by different gaze locations on the screen (Wang, 2011).

For fixed effects, Target Segment, Logged Target Word Frequency, Z-transformed Number of Homophones, and Participant's Mean Reaction Time were removed from the model during the model fitting and evaluating procedures, as they did not significantly improve the fit of the model (the variables that remained in the model are illustrated in Table 2). Word Duration and Baseline

Table 2. The summary of the model for L1 and L2 listeners in Go-NoGo task, showing the parametric coefficients and approximate significance of smooth terms in the model: estimated degrees of freedom (edf), reference degrees of freedom (Ref.df), F- and p-values for smooth terms

Parametric coefficients	Estimate	Std. Error	t-value	p-value
Intercept	36.038	9.100	3.960	< 0.0001
LanguageReduction: L2.Unreduced	-13.798	2.667	-5.174	< 0.0001
LanguageReduction: L1.Reduced	15.050	3.023	4.978	< 0.0001
LanguageReduction: L2.Reduced	-4.197	3.583	-1.171	0.241
Baseline Pupil Size	-0.023	0.002	-8.579	< 0.0001
Word Duration	32.823	13.945	2.354	< 0.05
Smooth terms	edf	Ref.df	F-value	p-value
Tensor product: Time and P-O Consistency Index: L1.Unreduced	8.790	8.972	68.716	< 0.0001
Tensor product: Time and P-O Consistency Index: L2.Unreduced	8.683	8.914	32.815	< 0.0001
Tensor product: Time and P-O Consistency Index: L1.Reduced	8.782	8.925	81.779	< 0.0001
Tensor product: Time and P-O Consistency Index: L2.Reduced	8.719	8.916	45.478	< 0.0001
Smooth: Trail Index	3.200	3.640	20.562	< 0.0001
Smooth: Gaze Coordinates X and Y	2.005	2.010	7.092	< 0.005
Random effect: ParIDConsis over time	1567.487	3047.000	1.374	< 0.0001
Random effect: ItemReduc over time	555.632	2256.000	0.586	< 0.0001

Pupil Size were refitted without a smooth function because their effects were linear and they can be included in the parametric coefficients. For random effects, we included two factor smooths: “ParIDConsis” (unique combination of Participant ID and P-O Consistency Index) for Time and “ItemReduc” (unique combination of Item (i.e., word) and Reduction) for Time. That is, we fitted separate factor smooths for each participant at each P-O Consistency Index to reflect participant-specific trends in the effect of P-O consistency over time, as well as for each item at each word form to take into account item-specific trends in the effect of reduction over time (See Wieling (2018) for an overview of random effects structures). After verifying the number of basis functions for the independent variables and interactions using the *gam.check* function (Please see Sóskuthy (2017) for an overview of basis functions), we included an AR-1 correlation parameter at the value of 0.980 in the model to address autocorrelation and also fitted the model with the scaled-*t* family in order for residuals to be normally distributed (Meulman et al., 2015; van Rij et al., 2019; Wieling, 2018).

Table 2 summarizes our final model, showing the parametric coefficients and approximate significance of smooth terms in the model: estimated degrees of freedom (edf), reference degrees of freedom (Ref.df), F- and p-values for smooth terms. The

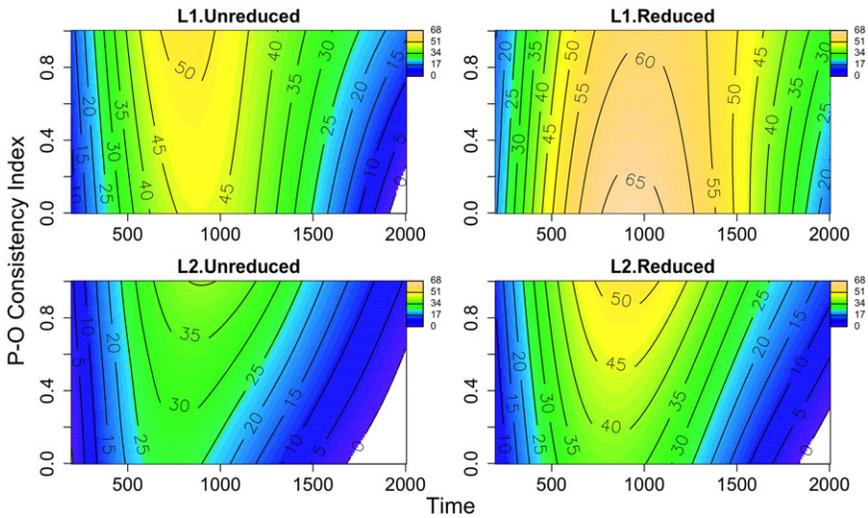


Figure 2. Contour plots of the interaction between the effect of Time (x-axis) and P-O Consistency Index (y-axis) by both Language and Reduction: L1.Unreduced (left top panel), L1.Reduced (right top panel), L2.Unreduced (bottom left panel), and L2.Reduced (bottom right panel). Shades of colours indicate the degree of pupil dilation: yellow: large dilation, green: medium dilation, blue: small dilation, and white: smaller than 0. The contour lines represent the pupil dilation values predicted by the model and their boundaries.

parametric coefficients indicate that (1) the overall pupil dilation is smaller for L2.Unreduced than for L1.Unreduced ($t = -5.174$, $p < 0.0001$), and (2) the overall pupil dilation is greater for L1.Reduced than for L1.Unreduced ($t = 4.978$, $p < 0.0001$). In addition, (3) the overall pupil dilation is smaller for larger baseline pupil sizes ($t = -8.579$, $p < 0.0001$), and (4) the overall pupil dilation becomes greater as word duration increases ($t = 2.354$, $p < 0.05$). The smooth terms reveal the significance of non-linear patterns associated with all the independent variables. We further discuss the summary of the final model along with visualization of the results.

The contour plots in Figure 2 illustrate the interaction between the effect of Time (x-axis) and P-O Consistency Index (y-axis) by both Language and Reduction: L1.Unreduced (left top panel), L1.Reduced (right top panel), L2.Unreduced (bottom left panel), and L2.Reduced (bottom right panel). Shades of colors indicate the degree of pupil dilation: yellow: large dilation, green: medium dilation, blue: small dilation, and white: smaller than 0. The contour lines represent the pupil dilation values predicted by the model and their boundaries. These plots demonstrate that (1) dilation peaks around 800 to 1000 ms for both word forms and for both listener groups. However, for L2 listeners the peaks come a little earlier for the low-consistency index than for the high consistency index. (2) Reduced forms elicit overall larger dilation and greater peak dilation than their unreduced counterparts for both groups, and (3) similar to Figure 1, the overall dilation is greater for L1 listeners than for L2 listeners. (4) All word forms and listener groups exhibit a gradient effect of P-O consistency. (5) There is greater pupil dilation as P-O consistency index decreases for L1.Reduced, while the others demonstrate that dilation becomes greater as P-O consistency index increases. That is, while inconsistent words induce

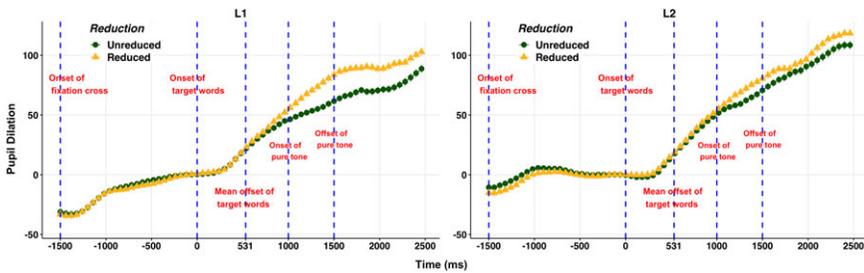


Figure 3. The grand average of pupillary dilation (y-axis) over time (x-axis) for reduced and unreduced word forms in the delayed naming task for L1 (left panel) and L2 (right panel) listeners. The vertical dot line at -1500 ms indicates the onset of the fixation cross, the line at 0 ms indicates the onset of stimuli, the line at 531 ms indicates the mean offset of stimuli, the line at 1000 ms indicates the onset of pure tone, and the line at 1500 ms displays the offset of pure tone.

greater dilation for L1.Reduced, consistent words elicit greater dilation for the others. The trend of the effect for L1.Reduced is in line with the results reported by Hino et al. (2017), but the trend of the effect for the others conflicts with the previous findings.

To further examine these aspects of the three-way interaction, we followed the procedure illustrated in Wieling (2018) and formally evaluated the difference between L1 and L2 listeners using binary difference smooths. We first decomposed the tensor product smooth interaction of P-O Consistency Index, LanguageReduction, and Time into separate parts using a *ti* constructor: the effect of Time by Reduction, the effect of P-O Consistency Index by Reduction, and the effect of the interaction between Time and P-O Consistency Index by Reduction. We then re-specified the model with a newly created binary variable representing if a given word is heard by L1 or L2 listeners (see Wieling (2018) for an overview of binary difference smooths and a *ti* constructor). The results demonstrate that the time course of dilation between L1 and L2 listeners differs for both reduced ($edf = 2.985$, $F = 4.721$, $p < 0.005$) and unreduced forms ($edf = 2.731$, $F = 3.968$, $p < 0.01$). Also, the trend of the P-O consistency effect between L1 and L2 listeners differs for both reduced ($edf = 2.605$, $F = 12.355$, $p < 0.0001$) and unreduced forms ($edf = 2.005$, $F = 11.318$, $p < 0.0001$). These results are in line with the description of the effects of (2) and (3) above. The trend of the interaction between Time and P-O consistency between L1 and L2 listeners does not differ for either forms (Reduced: $edf = 1.943$, $F = 1.863$, $p = 0.20831$; Unreduced: $edf = 1.590$, $F = 0.945$, $p = 0.31864$).

In sum, reduced forms elicit greater dilation than their unreduced counterparts for both L1 and L2 listeners, but the overall dilation appears to be greater for L1 listeners than for L2 listeners. The time course of dilation and the trend of the effect of consistency differ between the two listener groups for all the word forms.

Delayed naming task

Figure 3 displays the grand average of pupil dilation (y-axis) over time (x-axis) for reduced and unreduced word forms from -1500 ms to 2500 ms in the delayed naming task for L1 (left panel) and L2 (right panel) listeners. For both listener groups and for both word forms, pupil dilation increases as time progresses. Similar to the

Table 3. The summary of the model for L1 and L2 listeners in delayed naming task, showing the parametric coefficients and approximate significance of smooth terms in the model: estimated degrees of freedom (edf), reference degrees of freedom (Ref.df), F- and p-values for smooth terms

Parametric coefficients	Estimate	Std. error	t-value	p-value
Intercept	63.609	3.718	17.10	< 0.0001
LanguageReduction: L2.Unreduced	6.289	3.226	1.949	0.051
LanguageReduction: L1.Reduced	14.020	2.198	6.377	< 0.0001
LanguageReduction: L2.Reduced	14.199	3.342	4.248	< 0.0001
Baseline Pupil Size	-0.025	0.003	-7.637	< 0.0001
Smooth terms	edf	Ref.df	F-value	p-value
Tensor product: Time and P-O Consistency Index: L1.Unreduced	8.075	8.919	68.953	< 0.0001
Tensor product: Time and P-O Consistency Index: L2.Unreduced	7.006	7.727	68.573	< 0.0001
Tensor product: Time and P-O Consistency Index: L1.Reduced	8.434	8.752	95.534	< 0.0001
Tensor product: Time and P-O Consistency Index: L2.Reduced	8.016	8.521	77.446	< 0.0001
Smooth: Trail Index	3.456	3.818	44.542	< 0.0001
Smooth: Gaze Coordinates X and Y	8.688	8.978	84.702	< 0.0001
Random effect: ParIDConsis over time	1715.017	2917.000	1.660	< 0.0001
Random effect: ItemReduc over time	682.412	2257.000	1.218	< 0.0001

Go-NoGo task, for L1 listeners, the trend of pupil dilation over time appears to be comparable between the two forms, but reduced forms demonstrate greater peak dilation. The mean error rate for naming responses was lower than 1%. For L2 listeners, although the reduced form demonstrates a slightly greater dilation over time after 1000 ms, the difference between the two forms appears to be smaller than that of L1 listeners. Additionally, the dilation appears to be steeper and greater for L2 than for L1 listeners. The mean error rate for naming responses was 6.43 % (SD = 5.25).

We chose the time window from 200 ms to 2500 ms post-stimulus onset for this analysis. We used the same variables and procedures as earlier, but we did not include the Participant's Mean Reaction Time variable, as we did not measure the speed of naming for this task. For fixed effects, Target Segment, Logged Target Word Frequency, Word Duration, and Z-transformed Number of Homophones were removed from the model during the model fitting and evaluating procedures, as they did not significantly improve the fit of the model (the variables that remained in the model are illustrated in Table 3). Baseline Pupil Size was refitted without a smooth function because the effect was linear. As in the Go-NoGo task, two factor smooths were included for random effects: ParIDConsis (unique combination of Participant ID and P-O Consistency Index) for Time and

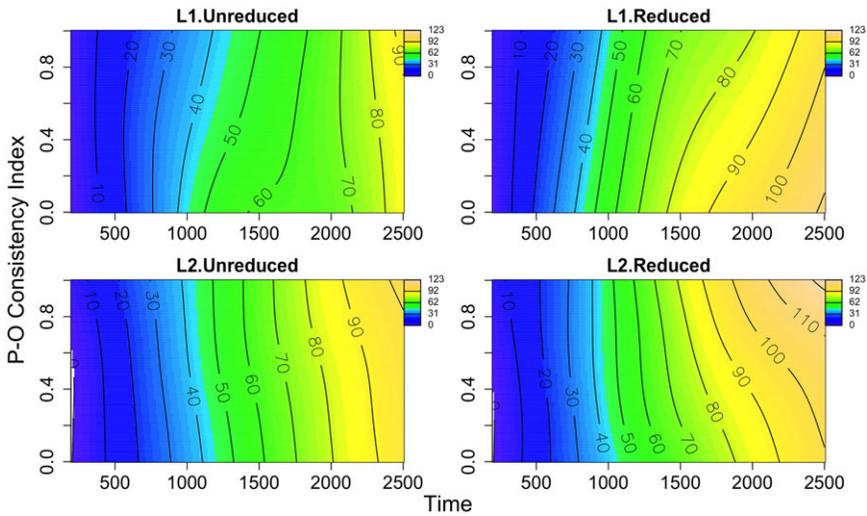


Figure 4. Contour plots of the interaction between the effect of Time (x-axis) and P-O Consistency Index (y-axis) by both Language and Reduction: L1.Unreduced (left top panel), L1.Reduced (right top panel), L2.Unreduced (bottom left panel), and L2.Reduced (bottom right panel). Shades of colours indicate the degree of pupil dilation: yellow: large dilation, green: medium dilation, blue: small dilation, and white: smaller than 0. The contour lines represent the pupil dilation values predicted by the model and their boundaries.

ItemReduc (unique combination of Item (i.e., word) and Reduction) for Time. We also included an AR-1 correlation parameter at the value of 0.985 and fitted the model with the scaled- t family. The summary of our final model is described in Table 3. The parametric coefficients indicate that (1) the overall pupil dilation is greater for L1.Reduced ($t = 6.377$, $p < 0.0001$) and L2.Reduced ($t = 4.248$, $p < 0.0001$) than for L1.Unreduced. In addition, (2) the overall pupil dilation is smaller for larger baseline pupil sizes ($t = -7.637$, $p < 0.0001$). Similarly, the smooth terms reveal the significance of non-linear patterns associated with all the independent variables.

As in the Go-NoGo task, Figure 4 demonstrates the interaction between the effect of Time and P-O Consistency Index by both Language and Reduction: L1.Unreduced (left top panel), L1.Reduced (right top panel), L2.Unreduced (bottom left panel), and L2.Reduced (bottom right panel). As indicated by the shades of colors and pupil dilation values of lines on the contour plots, we observe that: (1) for both word forms and both listener groups, the pupil dilates more greatly over time as we saw in Figure 3, and (2) reduced forms induce greater dilation than unreduced counterparts for both listener groups. (3) The consistency effect is larger for reduced forms than for unreduced counterparts for both listener groups, but (4) the direction of the effect of P-O consistency for reduced forms differs between L1 and L2 listeners. While pupil dilation becomes steeper and greater as P-O consistency index decreases for L1 listeners, L2 listeners demonstrate that dilation becomes steeper and greater as P-O consistency index increases. That is, while inconsistent words induce greater dilation for L1 listeners, consistent words elicit greater dilation for L2 listeners. As in the findings in the Go-NoGo task, the trend of the effect for

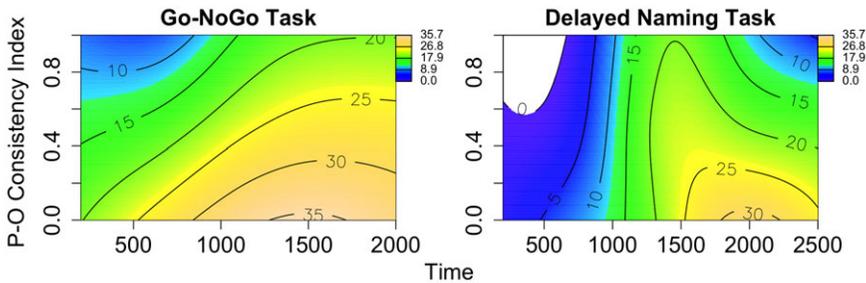


Figure 5. These contour plots demonstrate the difference in pupil dilation between reduced and unreduced forms over Time (x-axis) and P-O consistency Index (y-axis) in the Go-NoGo task (left panel) and delayed naming task (right panel) for L1 listeners (the difference was calculated by subtracting the pupil dilation value of unreduced forms from that of reduced forms). Shades of colours indicate the degree of difference in pupil dilation between the two forms: yellow: large difference, green: medium difference, blue: small difference, and white: smaller than 0. The contour lines represent the difference values in pupil dilation between the two forms and their boundaries.

L1 listeners is in line with the result of Hino et al. (2017), but the trend of the effect for L2 listeners conflicts with the previous findings.

As in the Go-NoGo task, we further examined these aspects of the three-way interaction by formally evaluating the difference between L1 and L2 listeners using binary difference smooths. The results demonstrate that only the trend of the P-O consistency effect between L1 and L2 listeners for reduced forms differs ($edf = 2.013$, $F = 3.150$, $p < 0.05$), which is in line with the discussion of (4). In short, for both listener groups, reduced forms appear to induce a greater dilation than unreduced counterparts, and the consistency effect is stronger for reduced forms than for unreduced forms. In addition, the effect of consistency differs between the two listener groups for reduced forms.

Overall, reduction and P-O consistency affect L1 and L2 listeners differently in both tasks. Further analyses were conducted focusing on the comparison between unreduced and reduced forms for L1 and L2 listeners separately, as well as their interaction with proficiency for L2 listeners. A full analysis and results of the L1 listeners, which mirror those reported in the previous section, are included in the supplementary material. A full analysis and results of the L2 listeners are also included in the supplementary material. The results of the L2 proficiency analysis are briefly discussed in the general discussion section, but it is important to note that the number of participants per group for the L2 proficiency analysis is small. Therefore, we need to be careful not to overly interpret the results.

Of particular interest is the interaction between the effect of orthography and the processing cost of reduced and unreduced forms. Previous research suggests that the processing advantage of unreduced forms could be due to a consistent relationship between an unreduced pronunciation and its orthographic form (Connine & Pinnow, 2006; Racine et al., 2014; Ranbom & Connine, 2007). However, our findings suggest that P-O consistency plays an important role in the processing of reduced forms as well. For example, Figure 5 illustrates differences in pupil dilation between the two forms over both Time (x-axis) and P-O consistency index (y-axis) in the Go-NoGo (left panel) and delayed naming task (right panel) for L1 listeners. The difference was

calculated by subtracting the pupil dilation value of unreduced forms from that of reduced forms. Similar to the earlier plots, shades of colors indicate the degree of difference in pupil dilation between the two forms: yellow: large difference, green: medium difference, blue: small difference, and white: smaller than 0. The contour lines represent the difference values in pupil dilation between the two forms and their boundaries. As shown by these plots, the difference in the processing cost between the two forms becomes smaller as the P-O consistency index increases, suggesting that the additional processing cost incurred by reduced forms is attenuated by the consistent P-O relationship. Consequently, the processing advantage of unreduced forms is greater for inconsistent words than for consistent ones. For example, in the Go-NoGo task, pupil dilation peaks approximately at 800 ms for unreduced forms and at 1000 ms for reduced forms (as shown in Figure 2). Within the time window (around 900 ms), the difference value is 30 at the lowest P-O consistency index but the value becomes half at the highest consistency index (left panel in Figure 5).

If the processing advantage of unreduced forms were due to a consistent relationship between an unreduced pronunciation and its orthographic form, as previous research suggested (Connine & Pinnow, 2006; Racine et al., 2014; Ranbom & Connine, 2007), we would not expect to observe the processing advantage for P-O inconsistent words. However, our findings suggest an advantage for unreduced forms for both consistent and inconsistent words in both tasks, suggesting that it is unlikely that consistency is the sole driving factor of the advantage. Instead, our result reveals a facilitatory effect for P-O consistent words in reduced forms, where the orthographic information that is consistent with its phonological form (phonology-to-orthography consistency) appears to reinforce the correct phonological representations. Our findings are in line with the notion raised by Mitterer and Reinisch (2015) where the P-O consistency effect should be re-conceptualized as facilitatory for consistent words instead of inhibitory for inconsistent words.

L2 advanced learners also demonstrate a similar pattern of the facilitatory effect for P-O consistent words in reduced forms, particularly in the Go-NoGo task (See the supplementary materials). However, the effect is not as pronounced as what has been found in L1 listeners. This could be due to either the qualitative difference between L1 and L2 listener's orthographic and phonological information in their lexical representations (Veivo & Järvikivi, 2013) or the proficiency difference between L1 and L2 listeners (i.e., L2 advanced listeners are not as proficient as L1 listeners).

General discussion

The present study investigated how the P-O consistency effect interacts with reduced pronunciations for L1 and L2 Japanese listeners. Go-NoGo and delayed naming tasks were employed, in combination with pupillometry, to compare the consistency effect between reduced and unreduced word forms over time. In particular, our study investigated whether the argument made by Mitterer and Reinisch (2015), that orthography does not play an important role in the perception of conversational reduced speech, holds for Japanese (a logographic language) with L1 and L2 listeners. Our predictions were that (1) if the orthographic effect does not play an

important role in the processing of reduced speech, we should observe an interaction between the effect of reduction and P-O consistency, indicating that P-O consistency affects reduced and unreduced pronunciations differently, and (2) the consistency effect would influence L1 and L2 listeners differently and the effect would be modulated by their proficiency for L2 listeners. We found that: (1) reduced forms elicit larger dilation than unreduced counterparts. (2) The P-O consistency effect emerges in both tasks, and the effect influences reduced and unreduced pronunciations for both L1 and L2 listeners differently, and (3) it is likely to be the case that the consistency effect varies depending on L2 proficiency (See the supplementary materials). Additionally, (4) there is a facilitatory effect for P-O consistent words instead of an inhibitory effect for P-O inconsistent words in reduced forms. These results are discussed in terms of their implications for how listeners process reduced speech and the role of the orthographic form in speech processing.

Orthographic consistency and reduced pronunciation

As in previous research (e.g., Brouwer et al., 2013; Ernestus et al., 2002; Janse et al., 2007; Tucker, 2011; van de Ven et al., 2011), we observed that unreduced forms are easier to process than their reduced counterparts in both tasks as evidenced by an overall smaller pupil dilation for unreduced forms. The difference between the two forms, however, appears to be smaller for L2 listeners. This is likely because some words are difficult to comprehend, particularly for basic and intermediate learners, regardless of word forms (reduced or not) due to the lack of lexical exposure. Importantly, we found a facilitatory effect for P-O consistent words in reduced forms, where the orthographic information that is consistent with its phonological form (phonology-to-orthography consistency) appears to be capitalized on to reinforce the correct phonological representations.

Crucially, this result is highly relevant to the discussion of the hypothesis that orthography plays a role in the processing of reduced speech. Our results demonstrate that acoustically reduced words require additional cognitive effort for processing, but orthography facilitates the processing and attenuates the additional processing cost, particularly for words that have a consistent P-O relationship. This result could extend the discussion of why Viebahn et al. (2018) did not find an effect of orthography in the processing of reduced pronunciation variants. In their study, the orthographic effect was measured by the presence or absence of orthographic <e> corresponding to the reduced pronunciation of /ə/ in novel French words. The researchers hypothesized that reduced pronunciations (without /ə/) that are spelled without <e> would be recognized faster than reduced pronunciations that are spelled with <e> because of the mismatch between the absence of /ə/ and the presence of <e>. They speculated that they did not find the orthographic effect because they did not consider the degree of spelling-to-pronunciation consistency between <e> and /ə/. Since <e> can be pronounced in multiple ways in French (e.g., /ə/, /ɛ/, /e/, or silent), the presence of <e> was not a reliable cue for the presence of /ə/. Accordingly, they predicted that they would find an effect of orthography if they examined an orthography that was highly consistent with its pronunciation.

While Viebahn et al. (2018) discussed their result from the spelling-to-pronunciation consistency point of view, we could further their discussion from the pronunciation-to-spelling consistency perspective. When stimuli are presented auditorily, phonological information is accessed before its orthographic information comes in (Rastle et al., 2011). It is, therefore, likely that consistency of phonology-to-orthography plays a more immediate and direct role than that of orthography-to-phonology (Ziegler et al., 2008). According to our results, the strong effect of orthography appears in the context of high P-O consistency for reduced pronunciations, meaning that in Viebahn et al. (2018), /ə/ should have a highly consistent P-O relationship with <e> to show the orthographic effect. However, since /ə/ is inconsistent with its orthographic form in French, they did not find an effect of orthography.

It is also possible to extend the discussion of our results from the phonological restructuring view. According to our findings, the degree of additional processing costs incurred by reduced pronunciations would be dependent on orthographic realizations. Specifically, how consistent the orthographic realization is with its phonological form (pronunciation-to-spelling consistency) modulates the extent to which the reduced pronunciation induces additional processing effort. In other words, the more consistent the orthographic realization is, the more restructuring it gives to its phonological representation. It then creates better specified and finer-grained phonological representations, which give a facilitatory effect when reduced consistent words are processed. This view is compatible with the discussion of the processing advantage of /nt/ (i.e., unreduced form) over the nasal flap [r] (i.e., reduced form) in English (Ranbom & Connine, 2007). Their hypothesis is that the [nt] processing advantage is due to orthographic forms that are consistent with their phonological form. From a frequency-based perspective, along with auditory input of the [nt] pronunciations, written input of <nt> orthography (that are consistent with its pronunciation of [nt]) may boost its frequency distribution. A nasal flap, however, does not receive the orthographic boost, as the [r] pronunciation and <nt> orthography are not consistent, meaning that its frequency distribution is based on the auditory domain solely. From the phonological restructuring view, we could argue that the orthographic form of [nt] is consistent with its phonological form. Therefore, it builds a finer phonological representation, which facilitates processing. An orthographic realization of nasal flap, however, is null, meaning that no phonological restructuring occurs that could build better specified phonological representations and facilitate its processing.

It is important to note that in our L1 results, the direction of the P-O consistency effect for unreduced forms in the Go-NoGo task was opposite to what has been found in previous studies—consistent words were more difficult to process than inconsistent words. This consistency effect in the reverse direction could be due to the relative time course of the effect and the task demands. According to Rastle et al. (2011), phonological representations are activated before the orthographic information comes in when the stimulus is a spoken word. When a given task demands less cognitive resources for its execution (e.g., comprehension of clearly articulated unreduced spoken words), the phonological activation can drive and complete the process before the orthographic information comes into effect. However, when the task requires a great amount of cognitive effort

(e.g., comprehension of acoustically reduced spoken words), additional processing time is needed. The extra cognitive effort and time provide an opportunity for orthographic information to be utilized to reinforce the correct phonological representation. In other words, the type of information that comes into play varies depending on task demands (Cutler et al., 2010; Norris et al., 2000), and orthographic information may be assigned less weight than phonological information as a processing cue because it is accessed indirectly, requiring an additional step (i.e., orthographic information is accessed via phonological representation). As a result, orthographic information does not play an important role when the processing effort is low. Accordingly, we could assume that the observed effect for unreduced forms in the Go-NoGo task is phonological because of its processing ease. The observed effect could be due to phonological neighbors. Ziegler et al. (2008) indicate that P-O consistency is naturally confounded with other lexical variables. We found that our low P-O consistent words tend to have a higher number of phonological neighbors. Cross-linguistic studies show that while an inhibitory effect of phonological neighborhood density has been found in English and French (i.e., high phonological neighborhood density leads to slower reaction time (e.g., Luce & Pisoni, 1998; Sommers, 1996; Vitevitch & Luce, 1999; Ziegler & Muneaux, 2007; Ziegler et al., 2003) and larger pupillary response (McLaughlin et al., 2022), a facilitatory effect has also been found in other languages such as Spanish and Russian (e.g., Arutiunian & Lopukhina, 2020; Vitevitch & Rodríguez, 2005).

For Japanese, work by Yoneyama (2002) shows that high phonological neighborhood density facilitates the recognition of Japanese words. In line with Yoneyama's result, the facilitatory effect of high phonological neighborhood density might have completed the process of unreduced words before the inhibitory effect of low P-O consistency would come into effect in the Go-NoGo task. This discussion also applies to the result of L2 listeners (See the supplementary materials). Additionally, the reason that we did not observe the reverse direction of consistency effect for unreduced forms in the delayed naming task could be because the overall cognitive demands were greater for the delayed naming task than for the Go-NoGo task.

Furthermore, as for the delayed naming task, our L1 findings appear to indicate a modest effect of P-O consistency. Our results are in line with the findings of previous studies, and they shed light on the interaction between the effect of the time course and reduction. The modest consistency effect emerges around 1200 to 1500 ms in unreduced forms, and the consistency effect appears to be stronger and it arises from 1500 to 2500 ms in reduced forms. Even when we consider the slowness of pupillary response (200 to 300 ms after a target cognitive event (Beatty, 1982)), the effect arises after the onset of the naming signal (shown by Figure 3), suggesting that the effect is likely involved in postlexical access stages (i.e., speech planning and production) rather than the perception stage. As a result, similar to the analysis of Bürki et al. (2012), in future work we should analyze participant's productions and compare it to our experimental stimuli to investigate how participants adjust their production of words as they hear either reduced or unreduced word forms, and how such adjustments interact with the P-O consistency of words.

It is also noteworthy that our L2 listeners are native speakers of English, meaning that the orthographic system between their L1 (English) and L2 (Japanese) differs substantially (alphabetic vs. logographic). According to the psycholinguistic grain size theory (Ziegler & Goswami, 2005), there is a different degree of phonological awareness for a different grain size of unit. For English, a rhyme is a salient unit of mapping between phonology and orthography; however for Japanese, a mapping between phonology and orthography is at the logographic character and syllable levels. We need to be careful not to overly interpret our result because it is possible that the substantially different levels of mapping processes could mean that even for advanced learners, they may not be able to access and utilize the information of P-O consistency during spoken word processing.

Orthographic consistency, reduced pronunciation, and task type

One of the issues raised by Mitterer and Reinisch (2015) is the relationship between speech type and task. As indicated by a number of studies (e.g., Cutler et al., 2010; Ranbom & Connine, 2007; Viebahn & Luce, 2020), an effect of orthography (e.g., P-O consistency) and speech style (e.g., reduced or unreduced word forms) appears to interact with a type of task that participants perform. Mitterer and Reinisch (2015) noted that the contrast between the two experiments in their study confounds speech type with task. That is, they found an orthographic effect with an explicit task (pronunciation judgment task) with careful speech (unreduced forms with limited contextual information) but they did not find the effect with an implicit task (visual world paradigm) with casual speech (reduced forms with conversational speech context), suggesting that further research is needed to investigate the effect in an implicit task with careful speech and in an explicit task with casual speech. Our results suggest that it is unlikely that orthography plays an important role in the comprehension of careful speech (unreduced forms in isolation) in an implicit task (delayed naming and Go-NoGo task). Additionally, in contrast to the results of Mitterer and Reinisch (2015), our findings indicate that the effect of P-O consistency emerges in implicit tasks (both in the Go-NoGo task and delayed naming task) with casual speech (reduced forms in isolation). This contrasting result could be because of the difference in the degree of “conversational likeness.” While reduced forms were presented with informal sentences including discourse markers and contractions in Mitterer and Reinisch (2015), stimuli were presented in isolation in our study. Mitterer and Reinisch (2015) discussed possibilities that presentation of stimuli in isolation could facilitate the activation of orthography. This notion is in line with our earlier discussion of the relative time course of orthographic effects and task demands. When contextual information is provided together with stimuli (e.g., sentence or discourse), linguistic cues in the contextual information can drive and complete the processing, but when no contextual information is provided (e.g., stimuli in isolation), additional cognitive effort and processing time is required, and it provides an opportunity for low weighted processing cues, such as orthography, to come into effect. Our results further provide evidence for the role of contextual information (with what information stimuli are presented) in the processing of reduced speech and its relationship with its orthographic form.

Methodological considerations for orthographic consistency and reduced pronunciation effect

Importantly, P-O consistency has been treated as a binary variable (consistent vs. inconsistent words) although research (Jared et al., 1990; Ziegler et al., 2004) indicates that the degree of (in)consistency is crucial. Work by Ziegler et al. (1996); Ziegler et al. (1997) and Berndt et al. (1987) demonstrate a statistical analysis of bidirectional inconsistency between spelling and pronunciation in monosyllabic French and English words. In addition to the categorization of P-O and O-P consistent words, they also provide prior and conditional probabilities. As a result, the conditional probabilities are connected to Jared et al.'s notion of a "friend word" and an "enemy word" and how the size of consistency effect is related to the summed frequency of these friend and enemy words. Work by Hino et al. (2017, 2011) and our study adapted this notion to compute the consistency. Our research is one of the few studies that follows Ziegler et al. (1996, 1997) and employed a P-O consistency scale to examine how the degree of the consistency interacts with the amount of additional processing effort induced by reduced pronunciations during spoken word processing.

Additionally, our results extend the discussion in Viebahn and Luce (2020) and shed light on the role of methodology. The researchers demonstrated that the processing difficulty of reduced pronunciations, which was found in a lexical decision task, disappeared in a shadowing task. They argue that meta-linguistic decision-making processes in lexical decisions amplify the effect of processing difficulty for reduced pronunciations. We however found a reduction effect in both delayed naming and Go-NoGo tasks. Although our tasks do not require linguistically motivated decision-making for target words, we found the effect. This contrasting result could be due to the measures employed. While we measured pupil dilation, they measured accuracy rates and response latencies. As Goldinger and Papesh (2012) reported, pupillary responses can be used to examine lexical processes and reveal differences in cognitive effort even when behavioral measures exhibit equivalent performance.

Conclusion

Drawing on the pupillometry technique, the current study suggests that orthographic realization matters in the processing of reduced pronunciations for L1 and L2 advanced listeners. Specifically, how consistent the orthographic realization is with its phonological form (phonology-to-orthography consistency) modulates the extent to which the reduced pronunciation induces additional processing costs.

Replication package. The supplementary materials, data, stimulus lists, and analysis code for this study are available at <https://osf.io/je3t4/>.

Acknowledgements. This research was supported in part by a Social Sciences and Humanities Research Council (SSHRC) Partnership Grant, "Words in the World," 895-2016-1008. The authors would like to thank the editor, three anonymous reviewers, Dr. Koji Miwa, and the Graduate School of Humanities at Nagoya University for assistance with data collection, and Dr. Matthew C. Kelley for assisting with the P-O consistency script.

Competing interest. We declare that we have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Notes

- 1 We are aware that the acquisition of a third or more language is an area of active research in multilingualism and language acquisition (see Puig-Mayenco et al. (2020) for a review of studies in L3 acquisition). However, the knowledge of additional languages beyond Japanese and English is beyond the scope of our study. Therefore, an L2 here refers to a second language as well as to any other non-native language in the process of being acquired (De Angelis, 2007).
- 2 More than 50% of their trials contained excessive blinks and artifacts (more than 30% of data points in these trials were eye blinks or artifacts).
- 3 The relatively high percentage of data removed due to the excessive blinks was possibly caused by the long length of the experiment period (45 minutes for each task). Additional analyses were performed to verify this preprocessing procedure. The authors observed a comparable trend in the dataset regardless of preprocessing criteria. See the supplementary material at <https://osf.io/jc3t4/>.
- 4 The reason that the overall pupil dilation is smaller for L2 participants than L1 participants could be due to the task. As illustrated in the present study section, our Go-NoGo task did not require the extraction of word meaning or the explicit analysis of phonological structure or lexical status, meaning that participants did not need to make linguistically derived decisions on the target items (passive listening), meaning that linguistic related analyses that L2 learners usually struggle with (requiring greater cognitive demands) are not included in this task. In addition, our results show that L2 participants, specifically basic and intermediate learners, are not as sensitive to P-O consistency and reduced pronunciation as L1 participants are. That is, low-consistency words and reduced pronunciations do not induce additional cognitive demands for L2 participants when compared to L1 participants. Therefore, the overall dilation was lower for L2 listeners.

References

- Arai, T. (1999). A case study of spontaneous speech in Japanese. *14th International Congress of Phonetic Sciences (ICPhS XIV)*, 1, 615–618.
- Arai, T., Warner, N., & Greenberg, S. (2007). Analysis of spontaneous Japanese in a multi-language telephone-speech corpus. *Acoustical Science and Technology*, 28(1), 46–48. doi: [10.1250/ast.28.46](https://doi.org/10.1250/ast.28.46)
- Arutiunian, V., & Lopukhina, A. (2020). The effects of phonological neighborhood density in childhood word production and recognition in Russian are opposite to English. *Journal of Child Language*, 47(6), 1244–1262. doi: [10.1017/S0305000920000112](https://doi.org/10.1017/S0305000920000112)
- Aston-Jones, G., Rajkowski, J., & Cohen, J. (1999). Role of locus coeruleus in attention and behavioral flexibility. *Biological Psychiatry*, 46(9), 1309–1320. doi: [10.1016/S0006-3223\(99\)00140-7](https://doi.org/10.1016/S0006-3223(99)00140-7)
- Baayen, H., Vasishth, S., Bates, D. M., & Kliegl, R. (2017). The cave of shadows: Addressing the human factor with generalized additive mixed models. *Journal of Memory and Language*, 94, 206–234. doi: [10.1093/jnci/85.5.365](https://doi.org/10.1093/jnci/85.5.365)
- Bassetti, B. (2006). Orthographic input and phonological representations in learners of Chinese as a foreign language. *Written Language and Literacy*, 9(1), 95–114. doi: doi.org/10.1075/wll.9.1.07bas
- Bates, E., & Liu, H. (1996). Cued shadowing. *Language and Cognitive Processes*, 11(6), 577–582. doi: [10.1080/016909696386962](https://doi.org/10.1080/016909696386962)
- Beatty, J. (1982). Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological Bulletin*, 91(2), 276–292. doi: [10.1037/0033-2909.91.2.276](https://doi.org/10.1037/0033-2909.91.2.276)
- Beatty, J., & Lucero-Wagoner, B. (2000). The pupillary system. In J. T. Cacioppo, L. Tassinary, & G. Berntson (Eds.), *Handbook of psychophysiology* (2nd ed., pp. 142–162). New York: Cambridge University Press.
- Berndt, R. S., Reggia, J. A., & Mitchum, C. C. (1987). Empirically derived probabilities for grapheme-to-phoneme correspondences in English. *Behavior Research Methods, Instruments, and Computers*, 19, 1–9.
- Brouwer, S., Mitterer, H., & Huettig, F. (2013). Discourse context and the recognition of reduced and canonical spoken words. *Applied Psycholinguistics*, 34(3), 519–539. doi: [10.1017/S0142716411000853](https://doi.org/10.1017/S0142716411000853)
- Bürki, A., Fougeron, C., Gendrot, C., & Frauenfelder, U. H. (2011). Phonetic reduction versus phonological deletion of French schwa: Some methodological issues. *Journal of Phonetics*, 39(3), 279–288. doi: [10.1016/j.wocn.2010.07.003](https://doi.org/10.1016/j.wocn.2010.07.003)

- Bürki, A., Spinelli, E., & Gaskell, M. G. (2012). A written word is worth a thousand spoken words: The influence of spelling on spoken-word production. *Journal of Memory and Language*, *67*, 449–467. doi: [10.1016/j.jml.2012.08.001](https://doi.org/10.1016/j.jml.2012.08.001)
- Bürki, A., Viebahn, M. C., Racine, I., Mabut, C., & Spinelli, E. (2018). Intrinsic advantage for canonical forms in spoken word recognition: myth or reality? *Language, Cognition and Neuroscience*, *33*(4), 494–511. doi: [10.1080/23273798.2017.1388412](https://doi.org/10.1080/23273798.2017.1388412)
- Chao, P.-C., Chen, W.-F., Zevin, J., & Lee, C.-Y. (2021). Neural correlates of phonology-to-orthography mapping consistency effects on Chinese spoken word recognition. *Brain and Language*, *219*, 104961. doi: [10.1016/j.bandl.2021.104961](https://doi.org/10.1016/j.bandl.2021.104961)
- Charoy, J., & Samuel, A. G. (2019). The effect of orthography on the recognition of pronunciation variants. *Journal of Experimental Psychology: Learning Memory and Cognition*, *46*(6), 1121–1145. doi: [10.1037/xlm0000781](https://doi.org/10.1037/xlm0000781)
- Chen, W.-F., Chao, P.-C., Chang, Y.-N., Hsu, C.-H., & Lee, C.-Y. (2016). Effects of orthographic consistency and homophone density on Chinese spoken word recognition. *Brain and Language*, *157–158*, 51–62. doi: [10.1016/j.bandl.2016.04.005](https://doi.org/10.1016/j.bandl.2016.04.005)
- Cohen, L., Jobert, A., Le Bihan, D., & Dehaene, S. (2004). Distinct unimodal and multimodal regions for word processing in the left temporal cortex. *NeuroImage*, *23*(4), 1256–1270. doi: [10.1016/j.neuroimage.2004.07.052](https://doi.org/10.1016/j.neuroimage.2004.07.052)
- Connine, C. M., & Pinnow, E. (2006). Phonological variation in spoken word recognition: Episodes and abstractions. *Linguistic Review*, *23*(3), 235–245. doi: [10.1515/TLR.2006.009](https://doi.org/10.1515/TLR.2006.009)
- Cutler, A., Treiman, R., & van Ooijen, B. (2010). Strategic deployment of orthographic knowledge in phoneme detection. *Language and Speech*, *53*(3), 307–320. doi: [10.1177/0023830910371445](https://doi.org/10.1177/0023830910371445)
- De Angelis, G. (2007). *Third or additional language acquisition* (Vol. 24). Multilingual Matters.
- Dehaene, S., Naccache, L., Cohen, L., Bihan, D. L., Mangin, J.-F., Poline, J.-B., & Rivière, D. (2001). Cerebral mechanisms of word masking and unconscious repetition priming. *Nature Neuroscience*, *4*(7), 752–758. doi: [10.1038/89551](https://doi.org/10.1038/89551)
- Dijkstra, T., Van Heuven, W. J. B., & Grainger, J. (1998). Simulating cross-language competition with the bilingual interactive activation model. *Psychologica Belgica*, *38*(3–4), 177–196.
- Ernestus, M., Baayen, H., & Schreuder, R. (2002). The recognition of reduced word forms. *Brain and Language*, *81*(1–3), 162–173. doi: [10.1006/brln.2001.2514](https://doi.org/10.1006/brln.2001.2514)
- Ernestus, M., & Warner, N. (2011). An introduction to reduced pronunciation variants. *Journal of Phonetics*, *39*(3), 253–260. doi: [10.1016/S0095-4470\(11\)00055-6](https://doi.org/10.1016/S0095-4470(11)00055-6)
- Escudero, P., Hayes-harb, R., & Mitterer, H. (2008). Novel second-language words and asymmetric lexical access. *Journal of Phonetics*, *36*(2), 345–360. doi: [10.1016/j.wocn.2007.11.002](https://doi.org/10.1016/j.wocn.2007.11.002)
- Escudero, P., & Wanrooij, K. (2010). The effect of L1 orthography on non-native vowel perception. *Language and Speech*, *53*(3), 343–365. doi: [10.1177/0023830910371447](https://doi.org/10.1177/0023830910371447)
- Frauenfelder, U. H., Segui, J., & Dijkstra, T. (1990). Lexical effects in phonemic processing: Facilitatory or inhibitory? *Journal of Experimental Psychology: Human Perception and Performance*, *16*(1). doi: [10.1037/0096-1523.16.1.77](https://doi.org/10.1037/0096-1523.16.1.77)
- Fushimi, T., Ijuin, M., Patterson, K., & Tatsumi, I. F. (1999). Consistency, frequency, and lexicality effects in naming Japanese Kanji. *Journal of Experimental Psychology: Human Perception and Performance*, *25*(2), 382.
- Geller, J., Still, M. L., & Morris, A. L. (2016). Eyes wide open: Pupil size as a proxy for inhibition in the masked-priming paradigm. *Memory and Cognition*, *44*(4), 554–564. doi: [10.3758/s13421-015-0577-4](https://doi.org/10.3758/s13421-015-0577-4)
- Goldinger, S. D., & Pappas, M. H. (2012). Pupil dilation reflects the creation and retrieval of memories. *Current Directions in Psychological Science*, *21*, 90–95. doi: [10.1177/0963721412436811](https://doi.org/10.1177/0963721412436811)
- Goldinger, S. D., Azuma, T., Abramson, M., & Jain, P. (1997). Open wide and say “Blah!”: Attentional dynamics of delayed naming. *Journal of Memory and Language*, *37*(2), 190–216. doi: [10.1006/jmla.1997.2518](https://doi.org/10.1006/jmla.1997.2518)
- Greenberg, S. (1999). Speaking in shorthand - a syllable-centric perspective for understanding pronunciation variation. *Speech Communication*, *29*(2), 159–176. doi: [10.1016/S0167-6393\(99\)00050-3](https://doi.org/10.1016/S0167-6393(99)00050-3)
- Guasch, M., Ferré, P., & Haro, J. (2017). Pupil dilation is sensitive to the cognate status of words: further evidence for non-selectivity in bilingual lexical access. *Bilingualism: Language and Cognition*, *20*(1), 49–54. doi: [10.1017/S1366728916000651](https://doi.org/10.1017/S1366728916000651)

- Han, J. I., Kim, J. Y., & Choi, T. H. (2020). The role of orthography in lexical processing of the phonological variants in second language. *Journal of Psycholinguistic Research*, *50*. doi: [10.1007/s10936-020-09725-4](https://doi.org/10.1007/s10936-020-09725-4)
- Haro, J., Guasch, M., Vallès, B., & Ferré, P. (2017). Is pupillary response a reliable index of word recognition? Evidence from a delayed lexical decision task. *Behavior Research Methods*, *49*(5), 1930–1938.
- Hastie, T., & Tibshirani, R. (1990). *Generalized additive models*. Chapman and Hall/CRC.
- Hayes-Harb, R., & Barrios, S. (2021). The influence of orthography in second language phonological acquisition. *Language Teaching*, *54*(3), 297–326. doi: [10.1017/S0261444820000658](https://doi.org/10.1017/S0261444820000658)
- Hayes-Harb, R., Brown, K., & Smith, B. L. (2018). Orthographic input and the acquisition of German final devoicing by native speakers of English. *Language and Speech*, *61*(4), 547–564. doi: [10.1177/0023830917710048](https://doi.org/10.1177/0023830917710048)
- Hayes-Harb, R., Nicol, J., & Barker, J. (2010). Learning the phonological forms of new words: Effects of orthographic and auditory input. *Language and Speech*, *53*(3), 367–381.
- Hino, Y., Kusunose, Y., & Lupker, S. J. (2017). Phonological-Orthographic consistency for Japanese words and its impact on visual and auditory word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, *43*(1), 126–146. doi: [10.1037/xhp0000281](https://doi.org/10.1037/xhp0000281)
- Hino, Y., Miyamura, S., & Lupker, S. J. (2011). The nature of orthographic-phonological and orthographic-semantic relationships for Japanese kana and kanji words. *Behavior Research Methods*, *43*(4), 1110–1151. doi: [10.3758/s13428-011-0101-0](https://doi.org/10.3758/s13428-011-0101-0)
- Hochmann, J.-R., & Papeo, L. (2014). The invariance problem in infancy: A pupillometry study. *Psychological Science*, *25*(11), 2038–2046.
- Hubert Lyall, I., & Järvikivi, J. (2021). Listener's personality traits predict changes in pupil size during auditory language comprehension. *Scientific Reports*, *11*(1), 5443. doi: [10.1038/s41598-021-84886-3](https://doi.org/10.1038/s41598-021-84886-3)
- Janse, E., Nootboom, S. G., & Quené, H. (2007). Coping with gradient forms of /t/-deletion and lexical ambiguity in spoken word recognition. *Language and Cognitive Processes*, *22*(2), 161–200. doi: [10.1080/0169096500371024](https://doi.org/10.1080/0169096500371024)
- Jared, D., McRae, K., & Seidenberg, M. S. (1990). The basis of consistency effects in word naming. *Journal of Memory and Language*, *29*(6), 687–715. doi: [10.1016/0749-596X\(90\)90044-Z](https://doi.org/10.1016/0749-596X(90)90044-Z)
- Kahneman, D., & Beatty, J. (1966). Pupil diameter and load on memory. *Science*, *154*(3756), 1583–1585.
- Karatekin, C., Couperus, J. W., & Marcus, D. J. (2004). Attention allocation in the dual-task paradigm as measured through behavioral and psychophysiological responses. *Psychophysiology*, *41*(2), 175–185. doi: [10.1111/j.1469-8986.2004.00147.x](https://doi.org/10.1111/j.1469-8986.2004.00147.x)
- Kaushanskaya, M., Blumenfeld, H. K., & Marian, V. (2019). The language experience and proficiency questionnaire (LEAP-Q): Ten years later. *Bilingualism: Language and Cognition*, 1–6. doi: [10.1017/S1366728919000038](https://doi.org/10.1017/S1366728919000038)
- Kramer, S. E., Lorens, A., Coninx, F., Zekveld, A. A., Piotrowska, A., & Skarzynski, H. (2013). Processing load during listening: The influence of task characteristics on the pupil response. *Language and Cognitive Processes*, *28*(4), 426–442. doi: [10.1080/01690965.2011.642267](https://doi.org/10.1080/01690965.2011.642267)
- Kryuchkova, T., Tucker, B. V., Wurm, L. H., & Baayen, R. H. (2012). Danger and usefulness are detected early in auditory lexical processing: Evidence from electroencephalography. *Brain and Language*, *122*(2), 81–91. doi: [10.1016/j.bandl.2012.05.005](https://doi.org/10.1016/j.bandl.2012.05.005)
- Kuchinke, L., Vo, M. L. H., Hofmann, M., & Jacobs, A. M. (2007). Pupillary responses during lexical decisions vary with word frequency but not emotional valence. *International Journal of Psychophysiology*, *65*(2), 132–140. doi: [10.1016/j.ijpsycho.2007.04.004](https://doi.org/10.1016/j.ijpsycho.2007.04.004)
- Laeng, B., Sirois, S., & Gredebäck, G. (2012). Pupillometry: A window to the preconscious? *Perspectives on Psychological Science*, *7*(1), 18–27. doi: [10.1177/1745691611427305](https://doi.org/10.1177/1745691611427305)
- Lõo, K., van Rij, J., Järvikivi, J., & Baayen, H. (2016). Individual differences in pupil dilation during naming task. In *Proceedings of the 38th annual conference of the cognitive science society* (pp. 550–555). Austin, TX: Cognitive Science Society.
- Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear and Hearing*, *19*, 1–36.
- Maekawa, K., Yamazaki, M., Ogiso, T., Maruyama, T., Ogura, H., Kashino, W., Koiso, H., Yamaguchi, M., Tanaka, M., & Den, Y. (2014). Balanced corpus of contemporary written Japanese. *Language Resources and Evaluation*, *48*(2), 345–371. doi: [10.1007/s10579-013-9261-0](https://doi.org/10.1007/s10579-013-9261-0)

- Marian, V., Blumenfeld, H., & Kaushanskaya, M. (2007). The language experience and proficiency questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech, Language, and Hearing Research*, *50*(4), 940–967. doi: [10.1044/1092-4388\(2007\)067](https://doi.org/10.1044/1092-4388(2007)067)
- Mathôt, S., Fabius, J., Heusden, E. V., & der Stigchel, S. V. (2018). Safe and sensible baseline correction of pupil-size data. *Behavior Research Methods*, *50*(1), 94–106. doi: [10.7287/peerj.preprints.2725v1](https://doi.org/10.7287/peerj.preprints.2725v1)
- Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing Type I error and power in linear mixed models. *Journal of Memory and Language*, *94*, 305–315. doi: [10.1016/j.jml.2017.01.001](https://doi.org/10.1016/j.jml.2017.01.001)
- McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: Part I. An account of basic findings. *Psychological Review*, *88*, 375–407. doi: [10.1192/s0368315x00238942](https://doi.org/10.1192/s0368315x00238942)
- McLaughlin, D. J., Zink, M. E., Gaunt, L., Spehar, B., Van Engen, K. J., Sommers, M. S., & Peelle, J. E. (2022). Pupillometry reveals cognitive demands of lexical competition during spoken word recognition in young and older adults. *Psychonomic Bulletin and Review*, *29*(1), 268–280.
- McLaughlin, D. J., Zink, M., Gaunt, L., Reilly, J., Sommers, M. S., Engen, K. V., & Peelle, J. E. (2022). *Give me a break! Unavoidable fatigue effects in cognitive pupillometry.* (<https://doi.org/10.31234/osf.io/z4quk>)
- McQueen, J. (1996). Word spotting. *Language and Cognitive Processes*, *11*(6), 695–699. doi: [10.1080/016909696387114](https://doi.org/10.1080/016909696387114)
- Metsala, J. L. (1997). An examination of word frequency and neighborhood density in the development of spoken-word recognition. *Memory and Cognition*, *25*(1), 47–56. doi: [10.3758/BF03197284](https://doi.org/10.3758/BF03197284)
- Metsala, J. L., & Walley, A. C. (1998). Spoken vocabulary growth and the segmental restructuring of lexical representations: Precursors to phonemic awareness and early reading ability. In J. L. Metsala & L. C. Ehri (Eds.), *Word recognition in beginning literacy* (pp. 89–120). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.
- Meulman, N., Wieling, M., Sprenger, S. A., Stowe, L. A., & Schmid, M. S. (2015). Age effects in L2 grammar processing as revealed by ERPs and how (not) to study them. *PLoS ONE*, *10*(12), 1–27. doi: [10.1371/journal.pone.0143328](https://doi.org/10.1371/journal.pone.0143328)
- Mitsugi, S. (2018). Proficiency influences orthographic activations during L2 spoken-word recognition. *International Journal of Bilingualism*, *22*(2), 199–214. doi: [10.1177/1367006916666388](https://doi.org/10.1177/1367006916666388)
- Mitterer, H., & Reinisch, E. (2015). Letters don't matter: No effect of orthography on the perception of conversational speech. *Journal of Memory and Language*, *85*, 116–134. doi: [10.1016/j.jml.2015.08.005](https://doi.org/10.1016/j.jml.2015.08.005)
- Morais, J., Cary, L., Alegria, J., & Bertelson, P. (1979). Does awareness of speech as a sequence of phones arise spontaneously? *Cognition*, *7*(4), 323–331. doi: [10.1016/0010-0277\(79\)90020-9](https://doi.org/10.1016/0010-0277(79)90020-9)
- Mukai, Y. (2020). *Production of reduced speech and the role of phonological-orthographic consistency* (Ph.D. thesis). The University of Alberta.
- Mukai, Y., Järvikivi, J., & Tucker, B. V. (2018). The effect of phonological-orthographic consistency on the processing of reduced and citation forms of Japanese words: Evidence from pupillometry. In E. Dmyterko (Ed.), *The 2018 annual conference of the Canadian linguistics association* (pp. 1–15).
- Mukai, Y., & Tucker, B. V. (2017). The phonetic reduction of nasals and voiced stops in Japanese across speech styles. In *Proceedings of the 31st General Meeting of the Phonetic Society of Japan, Tokyo: The Phonetic Society of Japan* (pp. 31–36).
- Muneaux, M., & Ziegler, J. (2004). Locus of orthographic effects in spoken word recognition: Novel insights from the neighbour generation task. *Language and Cognitive Processes*, *19*(5), 641–660. doi: [10.1080/01690960444000052](https://doi.org/10.1080/01690960444000052)
- Norris, D., McQueen, J., & Cutler, A. (2000). Merging information in speech processing: Feedback is never necessary. *Behavioral and Brain Sciences*, *23*(3), 299–370.
- Papesh, M. H., & Goldinger, S. D. (2012). Pupil-BLAH-metry: Cognitive effort in speech planning reflected by pupil dilation. *Attention, Perception, and Psychophysics*, *74*(4), 754–765. doi: [10.3758/s13414-011-0263-y](https://doi.org/10.3758/s13414-011-0263-y)
- Papesh, M. H., & Goldinger, S. D. (2015). Pupillometry and memory: External signals of metacognitive control. In G. H. Gendolla, M. Tops, & S. L. Koole (Eds.), *Handbook of Biobehavioral Approaches to Self-Regulation* (pp. 125–139). New York, NY: Springer.
- Pattamadilok, C., Knierim, I. N., Duncan, K. J. K., & Devlin, J. T. (2010). How does learning to read affect speech perception? *Journal of Neuroscience*, *30*(25), 8435–8444. doi: [10.1523/JNEUROSCI.5791-09.2010](https://doi.org/10.1523/JNEUROSCI.5791-09.2010)

- Pattamadilok, C., Kolinsky, R., Luksaneeyanawin, S., & Morais, J. (2008). Orthographic congruency effects in the suprasegmental domain: Evidence from Thai. *Quarterly Journal of Experimental Psychology*, *61*(10), 1515–1537. doi: [10.1080/17470210701587305](https://doi.org/10.1080/17470210701587305)
- Pattamadilok, C., Perre, L., Dufau, S., & Ziegler, J. C. (2009). On-line orthographic influences on spoken language in a semantic task. *Journal of Cognitive Neuroscience*, *21*(1), 169–179. doi: [10.1162/jocn.2009.21014](https://doi.org/10.1162/jocn.2009.21014)
- Peereman, R., Dufour, S., & Burt, J. S. (2009). Orthographic influences in spoken word recognition: The consistency effect in semantic and gender categorization tasks. *Psychonomic Bulletin and Review*, *16*(2), 363–368. doi: [10.3758/PBR.16.2.363](https://doi.org/10.3758/PBR.16.2.363)
- Perre, L., Bertrand, D., & Ziegler, J. C. (2011). Literacy affects spoken language in a non-linguistic task: An ERP study. *Frontiers in Psychology*, *2*, 1–8. doi: [10.3389/fpsyg.2011.00274](https://doi.org/10.3389/fpsyg.2011.00274)
- Perre, L., Pattamadilok, C., Montant, M., & Ziegler, J. C. (2009). Orthographic effects in spoken language: On-line activation or phonological restructuring? *Brain Research*, *1275*, 73–80. doi: [10.1016/j.brainres.2009.04.018](https://doi.org/10.1016/j.brainres.2009.04.018)
- Pitt, M. A. (2009). The strength and time course of lexical activation of pronunciation variants. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(3), 896–910. doi: [10.1037/a0013160](https://doi.org/10.1037/a0013160)
- Porretta, V., & Tucker, B. V. (2019). Eyes wide open: Pupillary response to a foreign accent varying in intelligibility. *Frontiers in Communication*, *4*, 1–12. doi: [10.3389/fcomm.2019.00008](https://doi.org/10.3389/fcomm.2019.00008)
- Porretta, V., Tremblay, A., & Bolger, P. (2017). Got experience? PMN amplitudes to foreign-accented speech modulated by listener experience. *Journal of Neurolinguistics*, *44*, 54–67. doi: [10.1016/j.jneuroling.2017.03.002](https://doi.org/10.1016/j.jneuroling.2017.03.002)
- Porretta, V., Tucker, B. V., & Järvikivi, J. (2016). The influence of gradient foreign accentedness and listener experience on word recognition. *Journal of Phonetics*, *58*, 1–21. doi: [10.1016/j.wocn.2016.05.006](https://doi.org/10.1016/j.wocn.2016.05.006)
- Puig-Mayenco, E., Alonso, J. G., & Rothman, J. (2020). A systematic review of transfer studies in third language acquisition. *Second Language Research*, *36*(1), 31–64. doi: [10.1177/0267658318809147](https://doi.org/10.1177/0267658318809147)
- Qu, Q., Cui, Z., & Damian, M. F. (2018). Orthographic effects in second-language spoken-word recognition. *Journal of Experimental Psychology: Learning Memory and Cognition*, *44*(8), 1325–1332. doi: [10.1037/xlm0000520](https://doi.org/10.1037/xlm0000520)
- R Development Core Team. (2018). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from <http://www.r-project.org/>
- Racine, I., Bürki, A., & Spinelli, E. (2014). The implication of spelling and frequency in the recognition of phonological variants: Evidence from pre-readers and readers. *Language, Cognition and Neuroscience*, *29*(7), 893–898. doi: [10.1080/01690965.2013.832784](https://doi.org/10.1080/01690965.2013.832784)
- Ranbom, L. J., & Connine, C. M. (2007). Lexical representation of phonological variation in spoken word recognition. *Journal of Memory and Language*, *57*(2), 273–298. doi: [10.1016/j.jml.2007.04.001](https://doi.org/10.1016/j.jml.2007.04.001)
- Rastle, K., McCormick, S. F., Bayliss, L., & Davis, C. J. (2011). Orthography influences the perception and production of speech. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(6), 1588–1594. doi: [10.1037/a0024833](https://doi.org/10.1037/a0024833)
- 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. doi: [10.3758/s13428-018-1134-4](https://doi.org/10.3758/s13428-018-1134-4)
- Saito, H., Masuda, H., & Kawakami, M. (1998). Form and sound similarity effects in kanji recognition. *Reading and Writing: An Interdisciplinary Journal*, *10*, 323–357. doi: [10.1023/A:1008093507932](https://doi.org/10.1023/A:1008093507932)
- Schmidtke, J. (2018). Pupillometry in linguistic research: An introduction and review for second language researchers. *Studies in Second Language Acquisition*, *40*(3), 529–549. doi: [10.1017/S0272263117000195](https://doi.org/10.1017/S0272263117000195)
- Seidenberg, M. S., & Tanenhaus, M. K. (1979). Orthographic effects on rhyme monitoring. *Journal of Experimental Psychology: Human Learning and Memory*, *5*(6), 546–554. doi: [10.1037/0278-7393.5.6.546](https://doi.org/10.1037/0278-7393.5.6.546)
- Showalter, C. E., & Hayes-Harb, R. (2015). Native English speakers learning Arabic: The influence of novel orthographic information on second language phonological acquisition. *Applied Psycholinguistics*, *36*(1), 23–42. doi: [10.1017/S0142716414000411](https://doi.org/10.1017/S0142716414000411)
- Sommers, M. S. (1996). The structural organization of the mental lexicon and its contribution to age-related declines in spoken-word recognition. *Psychology and Aging*, *11*(2), 333–341.
- Sóskuthy, M. (2017). Generalised additive mixed models for dynamic analysis in linguistics: A practical introduction. *arXiv*. doi: [10.48550/arXiv.1703.05339](https://doi.org/10.48550/arXiv.1703.05339)

- Stone, G. O., & Van Orden, G. C. (1994). Building a resonance framework for word recognition using design and system principles. *Journal of Experimental Psychology: Human Perception and Performance*, *20*(6), 1248–1268. doi: [10.1037/0096-1523.20.6.1248](https://doi.org/10.1037/0096-1523.20.6.1248)
- Stone, G. O., Vanhoy, M., & Van Orden, G. (1997). Perception is a two-way street: Feedforward and feedback phonology in visual word recognition. *Journal of Memory and Language*, *36*(3), 337–359. doi: [10.1006/jmla.1996.2487](https://doi.org/10.1006/jmla.1996.2487)
- Tamási, K., McKean, C., Gafos, A., Fritzsche, T., & Höhle, B. (2017). Pupillometry registers toddlers' sensitivity to degrees of mispronunciation. *Journal of Experimental Child Psychology*, *153*, 140–148. doi: <https://doi.org/10.1016/j.jecp.2016.07.014>
- Titone, C. M. C. D. (1996). Phoneme monitoring. *Language and Cognitive Processes*, *11*(6), 635–646. doi: [10.1080/016909696387042](https://doi.org/10.1080/016909696387042)
- Tucker, B. V. (2007). *Spoken word recognition of the reduced American English flap* (Ph.D. thesis). The University of Arizona.
- Tucker, B. V. (2011). The effect of reduction on the processing of flaps and /g/ in isolated words. *Journal of Phonetics*, *39*(3), 312–318. doi: [10.1016/j.wocn.2010.12.001](https://doi.org/10.1016/j.wocn.2010.12.001)
- van de Ven, M., Tucker, B. V., & Ernestus, M. (2011). Semantic context effects in the comprehension of reduced pronunciation variants. *Memory and Cognition*, *39*(7), 1301–1316. doi: [10.3758/s13421-011-0103-2](https://doi.org/10.3758/s13421-011-0103-2)
- van Rij, J., Hendriks, P., van Rijn, H., Baayen, R. H., & Wood, S. N. (2019). Analyzing the time course of pupillometric data. *Trends in Hearing*, *23*, 1–22. doi: [10.1177/2331216519832483](https://doi.org/10.1177/2331216519832483)
- van Rij, J., Vaci, N., Wurm, L. H., & Feldman, L. B. (2020). Alternative quantitative methods in psycholinguistics: Implications for theory and design. In V. Pirrelli, I. Plag, & W. U. Dressler (Eds.), *Word knowledge and word usage* (pp. 83–126). Berlin, Boston: De Gruyter Mouton.
- van Rij, J., Wieling, M., Baayen, R. H., & van Rijn, H. (2017). *itsadug: Interpreting time series and autocorrelated data using GAMMs*. (R package version 2.3)
- Vance, T. J. (2008). *The Sounds of Japanese*. Cambridge: Cambridge University Press.
- Veivo, O., & Järvikivi, J. (2013). Proficiency modulates early orthographic and phonological processing in L2 spoken word recognition. *Bilingualism-Language and Cognition*, *16*(4), 864–883. doi: [10.1017/S1366728912000600](https://doi.org/10.1017/S1366728912000600)
- Veivo, O., Järvikivi, J., Porretta, V., & Hyönä, J. (2016). Orthographic activation in L2 spoken word recognition depends on proficiency: Evidence from eye-tracking. *Frontiers in Psychology*, *7*. doi: [10.3389/fpsyg.2016.01120](https://doi.org/10.3389/fpsyg.2016.01120)
- Ventura, P., Morais, J., Pattamadilok, C., & Kolinsky, R. (2004). The locus of the orthographic consistency effect in auditory word recognition. *Language and Cognitive Processes*, *19*(1), 57–95. doi: [10.1080/01690960344000134](https://doi.org/10.1080/01690960344000134)
- Viebahn, M. C., & Luce, P. A. (2020). Where is the disadvantage for reduced pronunciation variants in spoken-word recognition? On the neglected role of the decision stage in the processing of word-form variation. *Language, Cognition and Neuroscience*, *35*(3), 339–359. doi: [10.1080/23273798.2019.1659987](https://doi.org/10.1080/23273798.2019.1659987)
- Viebahn, M. C., McQueen, J. M., Ernestus, M., Frauenfelder, U. H., & Bürki, A. (2018). How much does orthography influence the processing of reduced word forms? Evidence from novel-word learning about French schwa deletion. *Quarterly Journal of Experimental Psychology*, *71*(11), 2378–2394. doi: [10.1177/1747021817741859](https://doi.org/10.1177/1747021817741859)
- Vitevitch, M. S., & Luce, P. A. (1999). Probabilistic phonotactics and neighborhood activation in spoken word recognition. *Journal of Memory and Language*, *40*(3), 374–408.
- Vitevitch, M. S., & Rodriguez, E. (2005). Neighborhood density effects in spoken word recognition in Spanish. *Journal of Multilingual Communication Disorders*, *3*(1), 64–73. doi: [10.1080/14769670400027332](https://doi.org/10.1080/14769670400027332)
- Wang, J. (2011). Pupil dilation and eye tracking. In M. Schulte-Mecklenbeck, A. Kühberger, & R. Ranyard (Eds.), *A handbook of process tracing methods for decision research: A critical review and user's guide* (pp. 185–204). New York: Psychology Press.
- Warner, N., & Tucker, B. V. (2011). Phonetic variability of stops and flaps in spontaneous and careful speech. *The Journal of the Acoustical Society of America*, *130*(3), 1606. doi: [10.1121/1.3621306](https://doi.org/10.1121/1.3621306)
- Wieling, M. (2018). Analyzing dynamic phonetic data using generalized additive mixed modeling: A tutorial focusing on articulatory differences between L1 and L2 speakers of English. *Journal of Phonetics*, *70*, 86–116. doi: <https://doi.org/10.1016/j.wocn.2018.03.002>

- Wieling, M., Tomaschek, F., Arnold, D., Tiede, M., Bröker, F., Thiele, S., Wood, S. N., & Baayen, R. H. (2016). Investigating dialectal differences using articulatory. *Journal of Phonetics*, *59*, 122–143. doi: [10.1016/j.wocn.2016.09.004](https://doi.org/10.1016/j.wocn.2016.09.004)
- 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*, 233121651880086. doi: [10.1177/2331216518800869](https://doi.org/10.1177/2331216518800869)
- Wood, S. (2017). *mgcv: Mixed GAM computation vehicle with GCV/AIC/REML smoothness estimation*. (R package version 1.8-2.3)
- Wydell, T. N. (1998). What matters in kanji word naming: Consistency, regularity, or On/Kun-reading difference? *Reading and Writing*, *10*, 359–373. doi: [10.1023/A:1008083513500](https://doi.org/10.1023/A:1008083513500)
- Wydell, T. N., Butterworth, B., & Patterson, K. (1995). The inconsistency of consistency effects in reading: The case of Japanese Kanji. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(5), 1155–1168. doi: [10.1037/0278-7393.21.5.1155](https://doi.org/10.1037/0278-7393.21.5.1155)
- Wydell, T. N., Patterson, K. E., & Humphreys, G. W. (1993). Phonologically mediated access to meaning for Kanji: Is a rose still a rose in Japanese Kanji? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*(3), 491–514. doi: [10.1037/0278-7393.19.3.491](https://doi.org/10.1037/0278-7393.19.3.491)
- Yoneyama, K. (2002). *Phonological neighborhoods and phonetic similarity* (Ph.D. thesis). The Ohio State University.
- Zekveld, A. A., Heslenfeld, D. J., Johnsrude, I. S., Versfeld, N. J., & Kramer, S. E. (2014). The eye as a window to the listening brain: Neural correlates of pupil size as a measure of cognitive listening load. *NeuroImage*, *101*, 76–86. doi: [10.1016/j.neuroimage.2014.06.069](https://doi.org/10.1016/j.neuroimage.2014.06.069)
- 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. doi: [10.1097/AUD.0b013e3181d4f251](https://doi.org/10.1097/AUD.0b013e3181d4f251)
- Ziegler, J. C., & Ferrand, L. (1998). Orthography shapes the perception of speech: The consistency effect in auditory word recognition. *Psychonomic Bulletin and Review*, *5*(4), 683–689. doi: [10.3758/BF03208845](https://doi.org/10.3758/BF03208845)
- Ziegler, J. C., & Goswami, U. (2005). Reading acquisition, developmental dyslexia, and skilled reading across languages: A psycholinguistic grain size theory. *Psychological Bulletin*, *131*(1), 3–29. doi: [10.1037/0033-2909.131.1.3](https://doi.org/10.1037/0033-2909.131.1.3)
- Ziegler, J. C., & Muneaux, M. (2007). Orthographic facilitation and phonological inhibition in spoken word recognition: A developmental study. *Psychonomic Bulletin and Review*, *14*(1), 75–80.
- Ziegler, J. C., Ferrand, L., & Montant, M. (2004). Visual phonology: The effects of orthographic consistency on different auditory word recognition tasks. *Memory and Cognition*, *32*(5), 732–741. doi: [10.3758/BF03195863](https://doi.org/10.3758/BF03195863)
- Ziegler, J. C., Jacobs, A. M., & Stone, G. O. (1996). Statistical analysis of the bidirectional inconsistency of spelling and sound in French. *Behavior Research Methods, Instruments, and Computers*, *28*(4), 504–515. doi: [10.3758/BF03200539](https://doi.org/10.3758/BF03200539)
- Ziegler, J. C., Muneaux, M., & Grainger, J. (2003). Neighborhood effects in auditory word recognition: Phonological competition and orthographic facilitation. *Journal of Memory and Language*, *48*(4), 779–793.
- Ziegler, J. C., Petrova, A., & Ferrand, L. (2008). Feedback consistency effects in visual and auditory word recognition: Where do we stand after more than a decade? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(3), 643–661. doi: [10.1037/0278-7393.34.3.643](https://doi.org/10.1037/0278-7393.34.3.643)
- Ziegler, J. C., Stone, G. O., & Jacobs, A. M. (1997). What's the pronunciation for -ough and the spelling for /u/? A database for computing feedforward and feedback inconsistency in English. *Behavior Research Methods, Instruments, and Computers*, *29*(4), 600–618. doi: [10.3758/BF03210615](https://doi.org/10.3758/BF03210615)
- Zou, L., Desroches, A. S., Liu, Y., Xia, Z., & Shu, H. (2012). Orthographic facilitation in Chinese spoken word recognition: An ERP study. *Brain and Language*, *123*(3), 164–173. doi: [10.1016/j.bandl.2012.09.006](https://doi.org/10.1016/j.bandl.2012.09.006)

Cite this article: Mukai, Y., Järvi­kivi, J., and Tucker, BV. (2023). The role of phonology-to-orthography consistency in predicting the degree of pupil dilation induced in processing reduced and unreduced speech. *Applied Psycholinguistics* *44*, 784–815. <https://doi.org/10.1017/S0142716423000279>