The effect of phonological-orthographic consistency on the processing of reduced and citation forms of Japanese words: Evidence from pupillometry

Yoichi Mukai, Juhani Järvikivi and Benjamin V. Tucker
University of Alberta

1. Introduction

An inconsistent relationship between the pronunciation and spelling of words in various languages has been shown to affect the rate at which listeners recognize spoken words (Ziegler et al., 1997; Ziegler and Ferrand, 1998). In English, for example, the rhyme /-ʌk/ is consistently reflected in spelling as it has only one possible spelling ("-uck" as in "luck"), but the rhyme /-ip/ is not, because it has two possible spellings: "-eap" or "-eep" as in "leap" or "keep." As a result, the recognition of inconsistently spelled words, "leap" and "keep", is slower than that of consistently spelled words, "luck" (Ziegler and Ferrand, 1998). This effect, called *phonological-orthographic (P-O) consistency,* has been found in not only alphabetic languages but also logographic languages such as Chinese and Japanese. Unlike alphabetic languages, P-O consistency in logographic languages, has been measured based on homophone density, orthographic consistency, and frequency of phonological and orthographic neighbours (Chen et al., 2016; Hino et al., 2017).

These studies, it should be noted that, have focused exclusively on carefully pronounced words, namely, citation forms of words. This suggests that it is still unclear whether the P-O consistency effect applies only to the citation forms of words or generalizes to the reduced forms of words often found in conversational speech. In conversational speech, there are many instances of articulatory undershoot, leading to the reduction of acoustic signals. For example, in English, *yesterday* /jɛstɚdeɪ/ may be pronounced as [jɛʃeɪ] (Tucker, 2007), and in Japaneses, *daigaku* /daigaku/ 'university' could be produced as [daiakɯ] (Arai, 1999). This means that the pronunciation of many words in conversational speech is in fact inconsistent with the spelling due to this reduction. In the present study, we investigate how the P-O consistency effect interacts with phonetic reduction in a logographic language. Specifically, we are interested in the time-course of the P-O consistency effect with reduced forms of Japanese words as indexed by pupil dilation.

1.1 P-O consistency in Japanese

The degree to which a sound is consistently reflected in spelling impacts the recognition latency of spoken words, and if a sound is inconsistently reflected in spelling, the inconsistency delays recognition of spoken words (Ziegler et al., 2003). While there

Actes du congrès annuel de l'Association canadienne de linguistique 2018.
Proceedings of the 2018 annual conference of the Canadian Linguistic Association.

are many studies investigating the effect of P-O consistency on spoken word recognition, research on this effect among Japanese speakers is much more restricted. One experiment by Hino et al. (2017) tested the effect using the phonological-orthographic (P-O) consistency index. They calculated the index based on the phonological and orthographic neighbours of target words. They first identified the phonological neighbours of the target word and classified these neighbours into two types: "orthographic friend" and "orthographic enemy." If the phonological neighbour was also an orthographic neighbour of the target word, it was categorized as an orthographic friend; if not, it was categorized an orthographic enemy. Hino et al. (2017) defined phonological neighbours as words that differ by a single mora from the target word, and they defined orthographic neighbours as words that differ by a single character from the target word (Fushimi et al., 1999). For example, if the target word is [*gesuto* ゲスト /ge-su-to/ 'guest'], the phonological neighbours of the target word are [*kyasuto* キャスト /kja-su-to/ 'cast'] , [*tesuto* テスト /te-su-to/ 'test'], [*besuto* ベスト /be-su-to/ 'best'] and [*gesui* 下水 /ge-su-i/ 'sewer'] because all of these words differ by a single mora from the target word. [1] Of these words, *kyasuto*, *tesuto*, and *besuto* are orthographic friends as they differ by a single character from the target word, and *Gesui* is an orthographic enemy as more than one character is different from the target word (Hino et al., 2017). Note that Japanese uses multiple written scripts, one logographic script called 'Kanji', and two syllabic scripts called 'Hiragana' and 'Katakana', depending on a type of words, although the Kanji script is most commonly used.

After the classification of the phonological neighbours into orthographic friends and enemies, Hino et al. (2017) computed P-O consistency index as follows.

> **P-O consistency index** = (Target word frequency + Summed frequency of orthographic friends) / (Target word frequency + Summed frequency of orthographic friends and enemies)

The index ranges from 0 to 1, with 0 indicating low consistency and 1 indicating high consistency. If most of the phonological neighbours are also orthographic friends, the consistency index becomes higher, but if most of the phonological neighbors are orthographic enemies, the consistency index becomes lower. Hino et al. (2017) selected 48 logographic words based on the index, 24 high (mean consistency, 0.755) and 24 low (mean consistency, 0.039) consistency words, and conducted an auditory lexical decision task. They found an effect of P-O consistency; response latencies for low consistency words were slower than for high consistency words.

1.2 Phonetic reduction in Japanese

Most researchers have investigated the effect of P-O consistency using carefully pronounced words. If we consider conversational speech, however, another type of P-O

[1] Hyphens indicate mora boundaries.

inconsistency arises. Research shows that variation and reduction of acoustic signals are prevalent in conversational speech across languages (Ernestus and Warner, 2011; Barry and Andreeva, 2001). For example, Arai (1999) investigated pronunciation variants of Japanese words in telephone conversations and found a variety of reduction:

> Nasalized vowels before nasals: in case of the word /tenisu/, the nasal and following /i/ are deleted, and the /e/ is both nasalized and lengthened.
>
> /tenisu/ → [tẽːsɯ] 'tennis'
>
> Approximated voiced stops: the articulation of stop consonants is approximated due to the lack of full oral closure. In the extreme case, the consonants are deleted:
>
> /daigaku/ → [daiɣɑkɯ] or [daiakɯ] 'university'
> /b/ /d/ /g/ → [β] [ð] [ɣ] → [Ø]

In the present study, we focus on the reduction of consonants, specifically nasals and voiced stops, for two reasons. First, Mukai and Tucker (in preparation) have shown that speech style difference (read and conversational) affects the duration of consonants more than vowels in Japanese. Second, Arai (1999); Mukai and Tucker (2017) have demonstrated that nasals and voiced stops show the reduction of segments rather than the alternation in voicing.

1.3 The present study

The aim of this study is to compare the time-course of P-O consistency effect between reduced and citation forms of Japanese words as indicated by pupil dilation. The pupil has been shown to respond to physiological arousal during cognitive tasks (Beatty, 1982). The pupil dilates as cognitive tasks become more difficult, and peak dilation correlates with the amount of cognitive effort induced by tasks. (Zekveld et al., 2010). Pupil dilation has been utilized as index of cognitive load (Kahneman and Beatty, 1966) and applied to a variety of psycholinguistic studies, such as speech intelligibility (Zekveld and Kramer, 2014), speech planning (Papesh and Goldinger, 2012), word frequency (Kuchinke et al., 2007), and masked priming (Geller et al., 2016). Pupillometry offers a reliable method to examine allocations of cognitive resources imposed by different variables in speech processing (Laeng et al., 2012). For our experiment, pupillometry is particularly beneficial because it reflects the magnitude of cognitive effort over time in the absence of voluntary and conscious processes (Laeng et al., 2012) and it also reflects cognitive effort without the effect of task-specific strategies (Papesh and Goldinger, 2012).

Our research questions are twofold: (1) How does the P-O consistency effect interact with reduction in Japanese? (2) Do we see the same effect of P-O consistency between reduced and citation forms of Japanese words? If reduction influences P-O consistency, we would observe an interaction between the effect of reduction and P-O consistency. This suggests that reduction creates an additional mismatch between pronunciation and orthography and that the P-O consistency effect interacts with the actual pronunciation (the reduced form) of words. If reduction does not affect P-O consistency, we would not observe an interaction between the effect of reduction and P-O consistency. This means that reduction does not create an additional mismatch between pronunciation and orthography and that the P-O consistency effect interacts with the citation form of words.

2. Method

2.1 Participants

Thirty-eight native speakers of Japanese (female, $n = 16$) were recruited at Nagoya University and ranged in age from 18 to 25 years old (M = 19.7, SD = 1.69). All participants reported normal or corrected-to-normal vision and hearing.

2.2 Materials

We chose 226 disyllabic and digraphic words from the Balanced Corpus of Contemporary Written Japanese (Maekawa et al., 2014). These words consisted of a word-medial nasal or voiced stop. All words were recorded in both reduced and citation forms by a female native Japanese speaker, resulting in 452 total stimuli. We instructed the speaker to produce the words clearly for citation forms and casually (spontaneous speech like) for reduced forms. The speaker produced multiple tokens of both forms, and one of the researchers selected the most natural sounding tokens as stimuli. We then normalized the amplitude of the words. Table 1 shows acoustic properties of stimuli in both forms. In the table, a target segment (nasal or voiced stop) is represented as *TarSeg*, reduction (reduced or citation form) as *Reduc*, a word duration in milliseconds as *WorDur*, a target segment duration in milliseconds as *SegDur*, speech rate (the number of vowels per second) as *SRate*, a mean word pitch in Herz as *WordPit*, and an intensity difference in dB as *IntDiff*. We defined the intensity difference as the difference between the minimum intensity of the target segment to the averaged maximum intensity of surrounding segments (Tucker, 2011). The intensity difference measure was provided only for voiced stops because Mukai and Tucker (2017) suggest that the intensity difference might not be a reliable measure for the reduction of nasals. Overall, reduced forms displayed shorter duration, faster speech rate, lower mean pitch, and smaller intensity difference.

We created four stimulus lists, each of which contained 150 items: 115 target words, 30 non-target items and 5 practice words. The target words were counterbalanced across reduction, so that none of the lists consisted of the same word twice. Similar to Perre et al. (2011), we employed a 500-ms-long pure tone as non-target

SegType	Reduc	WorDur	SegDur	SRate	WorPit	IntDiff
Nasal	Citation	0.62	0.12	4.24	223.73	–
Nasal	Reduced	0.45	0.09	5.74	203.35	–
VoicedStop	Citation	0.61	0.05	5.08	222.07	15.02
VoicedStop	Reduced	0.46	0.04	6.73	204.33	13.40

Table 1: Mean values of acoustic properties of stimuli in reduced and citation forms.

items, and the ratio between the target and non-target trials was 70% and 30%. Participants' task was to identify a pure tone, so that they did not make any linguistically derived decision or that they did not respond to the target words (Perre et al., 2011).

2.3 Apparatus

We designed and controlled the experiment using SR Research Experiment Builder software. The movements of the right eye were tracked by the EyeLink II head-mounted eye-tracker (SR Research, Canada) in the pupil-only mode with a sampling rate of 250 Hz. We utilized the Etymotic Research insert ER1 earphones to present auditory stimuli and the 1024 × 768 resolution computer screen to present a fixation cross.

2.4 Procedure

Participants sat on a chair in a quiet room at a distance of approximately 60 to 80 cm from the computer screen. Luminance of the room was kept constant throughout the experiment. Participants were instructed to perform an auditory Go-NoGo task. In this task, participants respond to particular stimuli (Go) but they do not respond to a different set of stimuli (NoGo). In our experiment, participants looked at a fixation cross presented at the centre of the screen on a gray background for 1500 ms and heard either a target word or a pure tone (as they continued looking at the fixation cross). They then responded to either the pure tone by pressing a button on a Microsoft Side Winder game-pad or the target word by not pressing the button. The fixation cross disappeared 2000 ms after the onset of target words or after the button presses trigged by pure tones. In order to allow time for the pupil to settle back to baseline, a blank screen on a gray background remained for 4000 ms after the disappearance of the fixation cross.

We calibrated the eye-tracker prior to each session, as well as every 29 trials. We also ran drift-correction at the onset of every trial. Each session contained 150 trials, including 115 target words, 30 pure tones, and 5 practice items. The practice items were provided at the beginning of sessions to familiarize the participants with the task. The target words and pure tones were randomly assigned to each trial by the software. Participants took a brief break every 29 trials. The experiment lasted approximately 90 minutes.

3. Results

3.1 Preprocessing pupil size data

We visually inspected the range of pre and post blink artifacts and removed eye-blinks and their artifacts, 50 datapoints before and after the blinks. We then linearly interpolated the removed datapoints for each trial. When the initial or final datapoint in the trial was either the eye-blink or its artifact, the datapoint was 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.

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 stimuli to the onset of stimuli and performed standard baseline subtraction for each trial to quantify the degree of pupil dilation. We employed subtract baseline correction (absolute difference) rather than divisive baseline correction (proportional difference) because percentage measures are inflated when baseline pupil size is small (Beatty and Lucero-Wagoner, 2000; Mathôt, 2017). Relevant pupillary variables were computed on a trial-by-trial basis in the time window from the onset of stimuli to 2000 ms after the onset. We defined the peak dilation as the difference between the baseline pupil size and the maximum pupil size, as well as the peak latency as the time elapsed from the onset of stimuli to the offset of peak dilation.

The trials that contained excessive blinks and their artifacts (i.e. more than 30% of the trial) were excluded (12% of the data) and the participants who lost more than 50% of their trials due to the excessive blinks and their artifacts were also discarded (4 participants; 3.5% of the remaining data). Additional trials were excluded when the peak latency was shorter than 400 ms (13.2% of the remaining data), the peak dilation was smaller than 0 or bigger than 400 (3.3% of the remaining data), the baseline pupil size were more than 2 standard deviations apart from the mean baseline pupil size (3.6% of the remaining data) and the gaze location was more than 300 pixels apart from the fixation cross on the screen in either x- or y-axis (8.5% of the remaining data).

3.2 Data analysis

We applied generalized additive mixed modeling (GAMM) (Hastie and Tibshirani, 1990; Wood, 2006) to our data for two reasons. First, GAMM allows us to model non-linear relationships, as well as linear relationships, between a response variable and predictor variables. This is important because we expected pupil size to fluctuate over time. GAMM also can model two (or more) dimensional nonlinear interaction surfaces of continuous variables. Second reason that we applied GAMM is that it allows us to control serial dependency in time-series data, namely, autocorrelation. It considers the correlation between an observed value at time point t and an observed value at time point t+i (i $\geqq$1) in a time series (Baayen et al., 2017). By virtue of these functionalities, GAMM has been applied to model a variety of non-linear time series data, such as electromagnetic articulography data (Wieling et al., 2016), format trajectory data (Sóskuthy, 2017), visual world eye-tracking data (Veivo et al., 2016), and event-related potential data (Porretta et al., 2017).

The variables of interest were Pupil Dilation (in the standard arbitrary unit delivered by the eye tracking system) as a response variable, and P-O consistency index (0 - 1), Reduction (reduced or citation form), and Time (in milliseconds) as predictor variables. We also had Word Duration (in milliseconds), Reduced Segment (nasal or voiced stop), Baseline Pupil Size (same unit as Pupil Dilation), Pupil Gaze Coordinates X and Y (x- and y-axis eye gaze position on the screen in pixels), Trial Index (6 to 150), Logged Word Frequency, Logged Number of Phonological Neighbours, and Standardized Number of Homophones as control variables.

3.3 Summary of aggregated raw data

We inspected the aggregated raw pupil dilation data prior to fitting models. Figure 1 shows the grand average of pupillary responses over time for reduced and citation forms from -1500 ms (the onset of the fixation cross) to 2000 ms (the offset of the blank screen). Table 2 illustrates the means and standard deviations of peak dilation and latency for the two forms. 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.

Reduction	Peak dilation (SD)	Peak latency in ms (SD)
Citation	113.68 (76.21)	1166.72 (510.25)
Reduced	123.05 (78.08)	1176.74 (492.56)

Table 2: Means and standard deviations of peak dilation and latency for reduced and citation word forms.

3.4 Model fitting and evaluation

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 (2015), and Wieling (2018). We chose the time window from 200ms to 2000 ms post stimulus onset for our analysis, as reliable effects emerge slowly in pupillary response (200 to 300 ms) after a relevant cognitive event (Beatty, 1982).

Using a smooth function, Pupil Dilation was fitted as a function of P-O consistency index, Reduction and Time with Word Duration, Reduced Segment, Baseline Pupil Size, Pupil Gaze Coordinate X and Y, Trial Index, Logged Word Frequency,

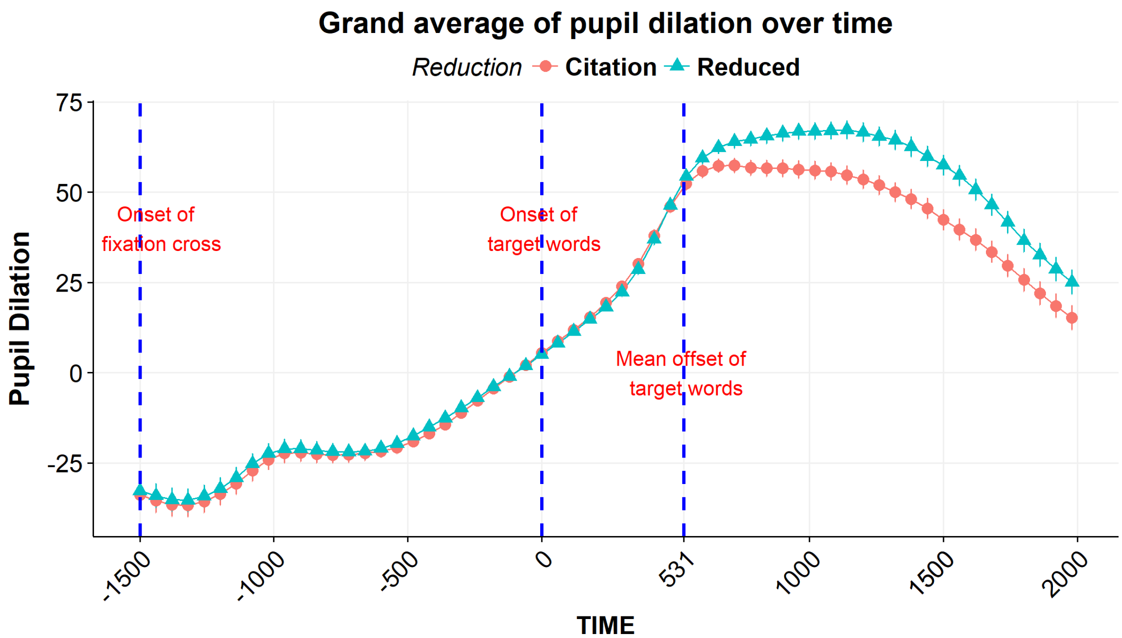

Figure 1: The grand average of pupillary responses over time for reduced and citation word forms. The vertical dot line at -1500 ms indicates the onset of the fixation cross, the line at 0 ms indicates the onset of stimuli and the line at 531 ms indicates the mean offset of stimuli.

Logged Number of Phonological Neighbours, and Standardized Number of Homophones. Furthermore, using a tensor product several types of interaction were included: Time, P-O consistency index and Reduction, Time and Logged Word Frequency, Time and Logged Number of Phonological Neighbours, Time and Standardized Number of Homophones, as well as Pupil Gaze Coordinate X and Y (See Wood, 2006 for an overview of a smooth function and tensor product). Inclusion of the three-way interaction between Time, Reduction and P-O consistency index allows us to examine the effect of each word form and their interaction over time. Inclusion of the two-way interactions between Time and the following variables (Logged Word Frequency, Logged Number of Phonological Neighbours, and Standardized Number of Homophones) reflects the possible difference in the effect of these variables over time, and the interaction between Pupil Gaze Coordinate X and Y captures the possible change in pupil size caused by different gaze locations on the screen (Wang, 2011).

We employed a backwards stepwise elimination procedure for fixed effects and a forward fitting procedure for random effects to fit the optimal model. We evaluated the contribution of input variables by χ^2 test of fREML scores using *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 input variables that were justified by the comparison ($p<.05$). Inclusion of interactions was also assessed by the fREML score comparison.

For fixed effects, we eliminated Logged Word Frequency, Logged Number of Phonological Neighbours, and Standardized Number of Homophones, including their interactions with Time. We also reduced the three-way interaction into two two-way interactions: Time and Reduction as well as P-O consistency index and Reduction. For random effects, while there are competing proposals regarding random-effects structures (Barr et al., 2013; Matuschek et al., 2017), we followed Wieling's approach (2018) and employed model selection to determine the optimal random-effects structure. 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 (See Wieling, 2018 for an overview of random-effects structures). That is, we fitted separate factor smooths for each participant at each P-O consistency index to reflect speaker-specific trends in the effect of P-O consistency, as well as for each item at each word form to take into account item-specific trends in the effect of reduction.

After verifying the number of basis functions for the predictor variables and interactions using *gam.check* function, [2] we included an AR-1 correlation parameter $\rho = 0.96$ to address autocorrelation and refitted the model with *scaled-t* family in order for residuals to be normally distributed (Meulman et al., 2015; Wieling, 2018). Table 3 illustrates the summary of the 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. In the summary, the parametric coefficients indicate the significant difference of overall pupil dilation between the reference level (citation form) and the reduced form. The smooth terms reveal the significance of non-linear patterns associated with the predictor variables except P-O consistency index. We further discuss the summary of the final model together with visualization of the results in the following section.

3.5 Statistical model

Figure 2 shows the time-course of pupil dilation for reduced and citation forms estimated by the model (left) and the comparison between the two forms (right). The plot on the left indicates that reduced forms elicit greater pupil dilation compared to citation forms. The plot on the right reveals the fact that the difference between the two forms are significant in the time window from 200 to 2000 ms. Figure 3 shows the degree of pupil dilation across the P-O consistency index for the two forms estimated by the model (left) and the comparison between the two (right). Similar to Figure 2, the plot on the left shows that reduced forms evoke larger pupil dilation

[2] The number of basis functions (knots) determines the degree of wiggliness of the estimated curve. Please see Wood (2006) or Sóskuthy (2017) for an overview of basis function.

Table 3: The summary of the 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.

Parametric coefficients	Estimate	Std.Error	t-value	p-value
Intercept	26.858	2.790	9.626	<2e-16***
Reduction:Reduced	22.744	4.072	5.585	2.34e-08***
Smooth terms	**edf**	**Ref.df**	**F-value**	**p-value**
s(Time):Citation	3.928	3.979	80.859	<2e-16***
s(Time):Reduced	3.927	3.978	126.731	<2e-16***
s(P-O Consistency):Citation	1.003	1.004	2.729	0.095458
s(P-O Consistency):Reduced	1.005	1.007	0.028	0.824102
s(BaselinePupilSize)	1.113	1.202	63.313	<2e-16***
s(WordDuration)	1.015	1.021	13.377	0.000523***
s(Trail Index)	5.305	6.351	25.323	<2e-16***
s(GazeX, GazeY)	6.395	7.743	15.435	<2e-16***
s(Time, ParIDConsis)	807.227	1773.000	1.481	<2e-16***
s(Time, ItemReduc)	555.410	2255.000	1.143	<2e-16***

than citation forms. The plot also reveals a trend, particularly citation forms, where the degree of pupil dilation becomes larger as P-O consistency increases. The plot on the right shows that the difference between the two forms are significant in the range of the P-O consistency index from 0 to 1. We need to, however, assess these differences formally whether the differences between the two forms are significant due to a constant difference (the overall degree of pupil dilation), a non-linear or linear difference (the trend of pupil dilation over time or the trend of pupil dilation across the P-O consistency index), or a combination of the constant and non-linear/linear differences. If the significance holds with the linear difference across the P-O consistency index, it means that there is an interaction between the effect of reduction and P-O consistency.

Following the procedure illustrated in Wieling (2018), we evaluated the differences using an ordered factor difference smooth (See Wieling, 2018 for an overview of an ordered factor difference smooth). The results demonstrate that the differences between the two forms are significant due to the overall degree of pupil dilation ($t=4.722$, $p<.001$) but not due to the trend of pupil dilation over time ($edf=3.48$, $F=1.28$, $p=0.27$) or the trend of pupil dilation across the P-O consistency index ($edf=1.01$, $F=2.15$, $p=0.12$). This suggests there is no interaction between the effect of reduction and P-O consistency; that is, reduction does not influence the P-O consistency effect. As a result, the P-O consistency effect does not interact with the actual pronunciation (reduced forms) of words. However, although there is no inter-

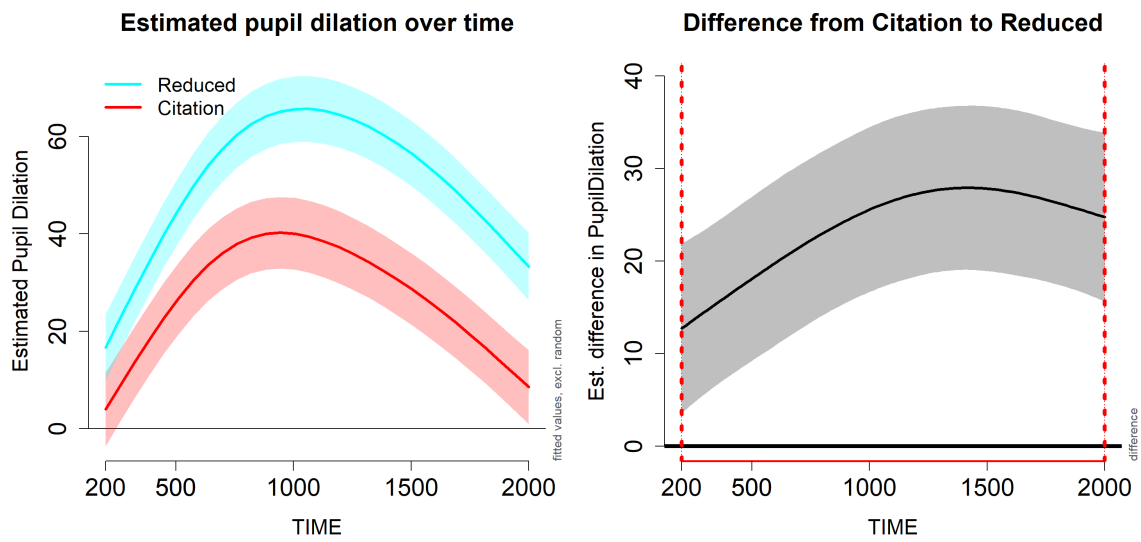

Figure 2: The time-course of pupil dilation for reduced and citation forms estimated by the model (left) and the comparison between the two forms (right). The vertical red dot lines in the comparison plot indicate that the difference between the two forms is significant between the two red lines.

action between the effect of reduction and P-O consistency, the relative size of P-O consistency effects visualized in Figure 3 reveals a trend in which reduced forms show a smaller P-O consistency effect as compared to citation forms. To better understand the trend, we need to fit models separately for the high and low P-O consistency index, as well as for reduced and citation forms. Furthermore, there is another trend found in Figure 3, in which the degree of pupil dilation becomes greater as P-O consistency increases. This result conflicts with what has been found by Hino et al. (2017). We need to study further about how to measure P-O consistency for Japanese words and how to interpret the results of the measurement, since we found that the way the P-O consistency index is calculated needs to be optimized and that the interpretation of the index does not seem to reflect what it is supposed to reflect.

4. Discussion

One possible interpretation of our results is that reduced forms do not impact the effect of P-O consistency, since reduced forms are connected to the citation forms in the mental lexicon and the reduced acoustic properties are restored based on the citation form during the recognition of spoken words (Kemps et al., 2004). This restoration process allows listeners to 'hear' the reduced acoustic information (Kemps et al., 2004), and the P-O consistency effect interacts with the restored citation forms. Finally, our results provide additional evidence that there is an influence of orthography

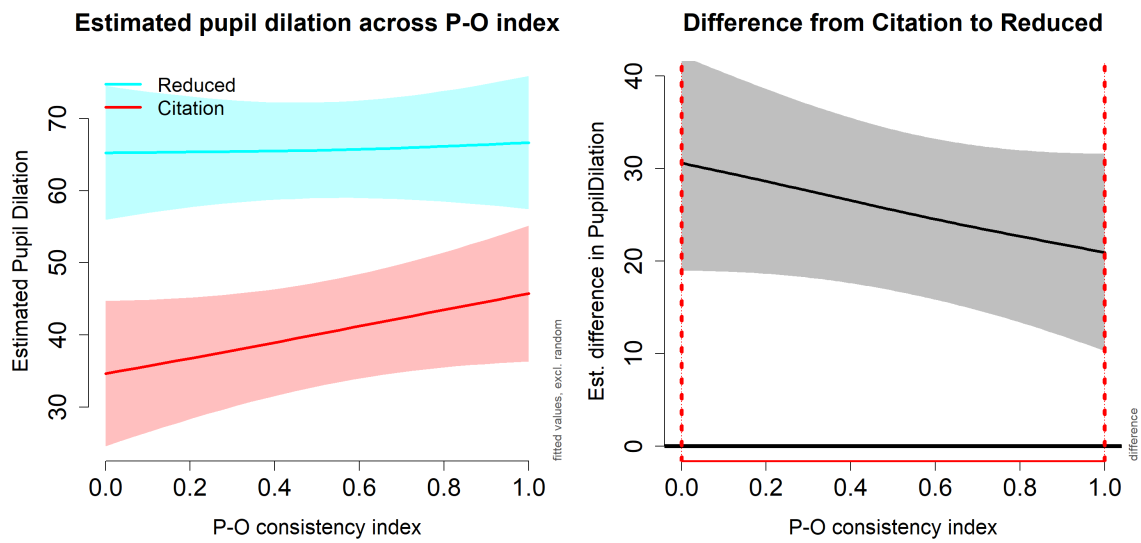

Figure 3: The degree of pupil dilation across the P-O consistency index for reduced and citation forms estimated by the model (left) and the comparison between the two forms (right). The vertical red dot lines in the comparison plot indicate that the difference between the two forms is significant between the two red lines.

in spoken word recognition. This orthographic influence challenges models of spoken word recognition, since it is unknown where the orthographic effect resides in the process of spoken word recognition. Taft (2011) argues that the orthographic information is integrated into abstract phonological representations; therefore, the orthographic effect resides in the domain of phonological processing. Furthermore, ERP studies have provided evidence for Taft's account in both alphabetic and logographic languages (Perre et al., 2009; Chen et al., 2016). However, there is one important issue to be addressed, particularly for Japanese. As discussed earlier, Japanese words can be written in multiple types of scripts, although the logographic script is most commonly used. In other words, it is unclear which script should be integrated into the phonological representations. Pylkkänen and Okano (2010) argue that, based on the results of visual word recognition experiments, multiple types of scripts are represented as part of the same representation if the scripts represent the same sound. In short, we need to investigate how multiple types of orthography are represented in the mind to further examine the effect of P-O consistency in Japanese.

5. Conclusion

The present study examined how the P-O consistency effect interacts with phonetic reduction in Japaneses by comparing the time-course of the effect between reduced and citation forms as indexed by pupillary response. Our results demonstrated that reduced forms elicit larger pupil dilation than citation forms and that reduction does not impact the effect of P-O consistency. We need further research for the nature of Japanese orthographic representation.

References

Arai, Takayuki. 1999. A case study of spontaneous speech in Japanese. *14th International Congress of Phonetic Sciences (ICPhS XIV)* 1: 615–618.

Baayen, Harald, Shravan Vasishth, D M Bates, and Reinhold Kliegl. 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.

Barr, Dale J., Roger Levy, Christoph Scheepers, and Harry J. Tily. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* 68(3): 255–278. doi:10.1016/j.jml.2012.11.001.

Barry, William and Bistra Andreeva. 2001. Cross-language similarities and differences in spontaneous speech patterns. *Journal of the International Phonetic Association* 31(1): 51–66. doi:10.1017/S0025100301001050.

Beatty, J and B Lucero-Wagoner. 2000. The pupillary system. In *Handbook of psychophysiology*, ed. J. T. Cacioppo, L.G. Tassinary, and G.G. Berntson, 2nd ed. New York: Cambridge University Press, 142–162.

Beatty, Jackson. 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.

Chen, Wei-fan, Pei-chun Chao, Ya-ning Chang, Chun-hsien Hsu, and Chia-ying Lee. 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.

Ernestus, Mirjam and Natasha Warner. 2011. An introduction to reduced pronunciation variants. *Journal of Phonetics* 39(3): 253–260. doi:10.1016/S0095-4470(11)00055-6.

Fushimi, Takao, Matsuo Ijuin, Karalyn Patterson, and Itaru F Tatsumi. 1999. Consistency, frequency, and lexicality effects in naming Japanese Kanji. *Journal of Experimental Psychology: Human Perception and Performance* 25(2): 382.

Geller, Jason, Mary L Still, and Alison L Morris. 2016. Eyes wide open : Pupil size as a proxy for inhibition in the masked-priming paradigm. *Memory & Cognition* : 554–564doi:10.3758/s13421-015-0577-4.

Hastie, T.J. and R.J. Tibshirani. 1990. *Generalized additive models*. Chapman and Hall/CRC.

Hino, Yasushi, Yuu Kusunose, and Stephen J Lupker. 2017. Phonological-Orthographic Consistency for Japanese Words and Its Impact on Visual and Auditory Word Recognition. *Journal of Experimental Psychology: Human Perception and Performance Human perception and performance* 43(1): 126–146.

Kahneman, D. and J. Beatty. 1966. Pupil Diameter and Load on Memory. *Science* .

Kemps, Rachèl, Mirjam Ernestus, Robert Schreuder, and Harald Baayen. 2004. Processing reduced word forms: The suffix restoration effect. *Brain and Language* 90(1-3): 117–127. doi:10.1016/S0093-934X(03)00425-5.

Kuchinke, Lars, Melissa L H Vo, Markus Hofmann, and Arthur M. Jacobs. 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.

Laeng, Bruno, Sylvain Sirois, and Gustaf Gredebäck. 2012. Pupillometry: A Window to the Preconscious? *Perspectives on Psychological Science* 7(1): 18–27. doi:10.1177/1745691611427305.

Maekawa, Kikuo, Makoto Yamazaki, Toshinobu Ogiso, Takehiko Maruyama, Hideki Ogura, Wakako Kashino, Hanae Koiso, Masaya Yamaguchi, Makiro Tanaka, and Yasuharu Den. 2014. Balanced corpus of contemporary written Japanese. *Language Resources and Evaluation* 48(2): 345–371. doi:10.1007/s10579-013-9261-0.

Mathôt, Sebastiaan. 2017. Safe and sensible baseline correction of pupil-size data. *PeerJ* (April): 1–25. doi:doi.org/10.7287/peerj.preprints.2725v1.

Matuschek, Hannes, Reinhold Kliegl, Shravan Vasishth, Harald Baayen, and Douglas Bates. 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.

Meulman, Nienke, Martijn Wieling, Simone A. Sprenger, Laurie A. Stowe, and Monika S. Schmid. 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.

Mukai, Yoichi and Benjamin V. Tucker. 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*.

Mukai, Yoichi and Benjamin V. Tucker. in preparation. Rhythm metrics and timing patterns of read and spontaneous speech: The case of Japanese, English, and L2 English .

Papesh, Megan H. and Stephen D. Goldinger. 2012. Pupil-BLAH-metry: Cognitive effort in speech planning reflected by pupil dilation. *Attention, Perception, & Psychophysics* 74(4): 754–765. doi:10.3758/s13414-011-0263-y.

Perre, Laetitia, Daisy Bertrand, and Johannes C. Ziegler. 2011. Literacy affects spoken language in a non-linguistic task: An ERP study. *Frontiers in Psychology* 2(OCT): 1–8. doi:10.3389/fpsyg.2011.00274.

Perre, Laetitia, Chotiga Pattamadilok, Marie Montant, and Johannes C. Ziegler. 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.

Porretta, Vincent, Antoine Tremblay, and Patrick Bolger. 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.

Pylkkänen, Liina and Kana Okano. 2010. The nature of abstract orthographic codes: Evidence from masked priming and magnetoencephalography. *PLoS ONE* 5(5). doi:10.1371/journal.pone.OOI0793.

R Development Core Team. 2018. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL `http://www.r-project.org/`.

Sóskuthy, Márton. 2017. Generalised additive mixed models for dynamic analysis in linguistics: a practical introduction URL `http://arxiv.org/abs/1703.05339`.

Taft, Marcus. 2011. Orthographic influences when processing spoken pseudowords: Theoretical implications. *Frontiers in Psychology* 2(JUN): 1–7. doi:10.3389/fpsyg.2011.00140.

Tucker, Benjamin V. 2007. *Spoken word recognition of the reduced American English flap*. Ph.D. thesis, The University of Arizona.

Tucker, Benjamin 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.

van Rij, Jacolien. 2015. Overview GAMM analysis of time series data Http:// www.sfs.uni-tuebingen.de/ jvanrij/Tutorial/GAMM.html. Accessed on 2/01/2018.

van Rij, Jacolien, Martijn Wieling, R. Harald Baayen, and Hedderik van Rijn. 2017. itsadug: Interpreting Time Series and Autocorrelated Data Using GAMMs. R package version 2.3.

Veivo, Outi, Juhani Järvikivi, Vincent Porretta, and Jukka Hyönä. 2016. Orthographic Activation in L2 Spoken Word Recognition Depends on Proficiency: Evidence from Eye-Tracking. *Frontiers in Psychology* 7(July). doi:10.3389/fpsyg.2016.01120.

Wieling, Martijn. 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* : 1–51.

Wieling, Martijn, Fabian Tomaschek, Denis Arnold, Mark Tiede, Franziska Bröker, Samuel Thiele, Simon N. Wood, and R. Harald Baayen. 2016. Investigating dialectal differences using articulography. *Journal of Phonetics* 59: 122–143. doi: 10.1016/j.wocn.2016.09.004.

Wood, Simon. 2017. mgcv: Mixed GAM computation vehicle with GCV/AIC/REML smoothness estimation. R package version 1.8-2.3.

Wood, Simon N. 2006. *Generalized additive models: An introduction with R*. Boca Raton, FL: Chapman & Hall/CRC Press. doi:10.1111/j.1541-0420.2007.00905_3.x.

Zekveld, Adriana A and Sophia E. Kramer. 2014. Cognitive processing load across a wide range of listening conditions: Insights from pupillometry. *Psychophysiology* 51: 277–284. doi:10.1111/psyp.12151.

Zekveld, Adriana A., Sophia E. Kramer, and Joost M. Festen. 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.

Ziegler, Johannes C. and Ludovic Ferrand. 1998. Orthography shapes the perception of speech: The consistency effect in auditory word recognition. *Psychonomic Bulletin & Review* 5(4): 683–689. doi:10.3758/BF03208845.

Ziegler, Johannes C., Mathilde Muneaux, and Jonathan Grainger. 2003. Neighborhood effects in auditory word recognition: Phonological competition and orthographic facilitation. *Journal of Memory and Language* 48(4): 779–793. doi:10.1016/S0749-596X(03) 00006-8.

Ziegler, Johannes C, Gregory O Stone, and Arthur M Jacobs. 1997. What's the pronunciation for − ough and the spelling for /u/? A database for computing feedforward and feed- back inconsistency in English. *Behavior Research Methods, Instruments, & Computers* 29(4): 600–618. doi:10.3758/BF03210615.