Production and perception of reduced speech and the role of phonological-orthographic consistency

by

Yoichi Mukai

A thesis submitted in partial fulfilment of the requirements for the degree of

Doctor of Philosophy

Department of Linguistics
University of Alberta

Examining committee:
Benjamin V. Tucker, Supervisor
Juhani Järvikivi, Supervisor
Anja Arnhold, Supervisory Committee
Trelani Chapman, Examiner
Timothy J. Vance, Examiner

Abstract

This dissertation examines the effect of orthography in the perception of spontaneous Japanese speech by investigating how phonetic reduction interacts with the effect of sound-to-spelling inconsistencies (i.e., phonological-orthographic (P-O) consistency effect) for L1 and L2 Japanese speakers. In order to accomplish this, we first conducted a corpus analysis to investigate the distribution and degree of phonetic reduction across various styles of speech. Specifically, we examined the effect of speech style on the realization of word-medial voiced stops and word-medial nasals across four different styles of speech (from spontaneous to read speech). The results indicated that the most spontaneous speech demonstrates greater reduction than the least spontaneous speech, and the acoustic results of reduction and their distributional patterns across speech styles for voiced stops are comparable between Japanese and English, and nasals also indicate comparable reduction patterns. We then conducted two pupillometry experiments (Go-NoGo and delayed naming tasks) to compare the time-course of the P-O consistency effect between reduced (spontaneous speech-like) and unreduced (read speech-like) word forms for L1 and L2 Japanese listeners. The two experiments provide evidence that the phonetic realization of Japanese words (reduced or unreduced) influences the consistency effect for both listeners during spoken word comprehension. Reduced word forms caused both L1 and L2 listeners to incur additional processing costs for comprehension, and both listeners exerted the P-O consistency information to increase efficacy in the processing of reduced words. While the processing cost of reduced forms was attenuated in consistent words for L1 listeners, the cost was attenuated in inconsistent words for L2 listeners. Within the L2 listeners, the high proficiency learners showed a clear P-O consistency effect that weakens as L2 pro-

ficiency decreases in the delayed naming task, but in the Go-NoGo task, the basic proficiency learners exhibited the clear consistency effect and it weakened as L2 proficiency increase. In summary, our findings suggest that the phonetic realization of Japanese words matters for the effect of P-O consistency and the consistency effect plays an important role in the processing cost of reduced forms.

Preface

This dissertation is original work by Yoichi Mukai. The research projects contained within this dissertation received research ethics approval from the University of Alberta Research Ethics Board 2, project name *The effect of orthography in spoken word recognition*, No. Pro00075224, 20 August 2019. The research conducted for this thesis was done in collaboration with Dr. Benjamin V. Tucker and Dr. Juhani Järvikivi.

The study in Chapter 2 was carried out with assistance from Dr. B.V. Tucker. I was responsible for data analysis and manuscript composition. Dr. B.V. Tucker assisted with concept formulation, data analysis and manuscript edits. Dr. J. Järvikivi and Dr. Anja Arnhold assisted with manuscript edits. Acquisition of the corpus data was supported by Prince Takamado Japan Centre for Teaching and Research at the University of Alberta.

The studies in Chapter 3 and 4 were carried out with assistance from Dr. B.V. Tucker and Dr. J. Järvikivi. I was responsible for experiment design, data collection, analysis, and manuscript composition. Dr. B.V. Tucker and Dr. J. Järvikivi assisted with concept formulation, experiment design, and data analysis and manuscript edits. Dr. A. Arnhold assisted with manuscript edits.

Recruitment for experiments in Chapter 3 and 4 was supported through a Social Sciences and Humanities Research Council (SSHRC) Partnership Grant, "Words in the World," 895- 2016-1008, to Dr. J. Järvikivi. No part of this dissertation has been previously published.

Acknowledgements

I would like to express my gratitude to all those who have guided and supported me throughout my studies. Specifically, I thank my supervisors, Drs. Benjamin V. Tucker and Juhani Järvikivi, for their guidance, support, patience, and mentorship. I also thank my supervisory committee member, Dr. Anja Arnhold, for valuable feedback on a draft of this dissertation, as well as examiners, Drs. Trelani Chapman and Timothy J. Vance, for insightful comments on this manuscript and during the dissertation defense.

I am also thankful for all the members in the Department of Linguistics at the University of Alberta, in particular, for the members of Alberta Phonetics Lab & Centre for Comparative Psycholinguistics.

Lastly, I wish to dedicate this dissertation to my parents and Stéphanie Martin who have been there for me throughout this PhD.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Most people spend many hours communicating using speech on a daily basis. In everyday interactions, our speech is produced in a casual manner, and such speech contains a high degree of variability. For example, Greenberg (1999) found in the Switchboard corpus (Godfrey, Holliman & McDaniel, 1992) that there were 117 different realizations of the word *that*, including [ðæ] being the most frequent and that there were 87 different realizations of the word *and*, including [æn], [ɛn], [ɪn], [ən], [n̩], [n] and [ænd]. Such variation is called phonetic reduction, where segments, syllables and even whole words can be altered and/or deleted (e.g., Ernestus, 2000; Johnson, 2004). Despite such variability, speakers and listeners seem to communicate with ease. It is an intriguing question to investigate how production of speech varies and how listeners interact with such variability.

Variability in production is not the only characteristic that complicates speech communication. Spoken language comprehension involves not only auditory information of utterances, but also visual information associated with the utterances. For example, the McGurk effect is a classic example demonstrating that auditory input and visual information from facial gestures interact during the comprehension of speech (McGurk & Macdonald, 1976), indicating that the comprehension of speech can be based on both auditory and visual information. One visual representation that is closely connected to speech and frequently utilized to communicate is written language, and research shows that orthographic knowledge influences the recognition of spoken language (Frauenfelder, Segui & Dijkstra, 1990; Jakimik, Cole & Rudnicky, 1985; Seidenberg & Tanenhaus, 1979; Ziegler & Ferrand, 1998).

In particular, how predictable or consistent the spelling of a given word is has been shown to affect its recognition.

In this dissertation, we compare the distribution and degree of phonetic reduction between casual and careful speech, and also examine how orthography affects the processing of casual speech compared to careful speech. This dissertation describes three studies. Chapter 2 describes a corpus analysis that examines phonetic variability of word-medial voiced stops and word-medial nasals in Japanese across different styles of speech, and Chapter 3 and 4 describe two pupillometry experiments that investigate how such variability interacts with the effect of pronunciation-to-spelling consistency during spoken word recognition in first (Chapter 3) and second (Chapter 4) language speakers. Importantly, we investigate Japanese which uses logographic scripts. Our general aims for this dissertation are twofold. The first is to understand how production of speech varies across different styles of speech and how listeners interact with such variability. The second is to understand how visual language information interacts with auditory language information during spoken language comprehension. Taking these two issues together is crucial to understanding how humans process spoken language using multiple sources of language information to cope with variability.

1.1 Orthography in spoken word recognition

The influence of orthography in spoken word recognition has been observed by a number of studies (Frauenfelder et al., 1990; Jakimik et al., 1985; Seidenberg & Tanenhaus, 1979; Ziegler & Ferrand, 1998), many of which concern the effect of phonological and orthographic (P-O) consistency. The inconsistent relationship between phonology (pronunciation) and orthography (spelling) has been shown to affect the recognition of spoken words, where inconsistent words are more difficult to process than consistent words (Pattamadilok, Perre, Dufau & Ziegler, 2009; Perre, Bertrand & Ziegler, 2011; Perre, Pattamadilok, Montant & Ziegler, 2009; Stone, Vanhoy & Van Orden, 1997; Ziegler & Ferrand, 1998; Ziegler, Ferrand & Montant, 2004; Ziegler, Van Orden & Jacobs, 1997). Consistent words are ones

where a single sound unit (e.g., rhyme) can be spelled in only one way, such as /-ʌk/ as in -uck in "luck", and inconsistent words are ones where a single sound unit can be spelled in multiple ways, such as /-ip/ as in -eap or -eep in "leap" or "keep".

However, this relationship between phonology and orthography is complicated by the fact that the realization of sound segments in spoken language is highly variable, especially in spontaneous, conversational speech (Warner & Tucker, 2011). This variability is often the result of phonetic reduction, realized as shortening, deletion, and/or incomplete articulation of segments (Ernestus & Warner, 2011; Greenberg, 1999; Warner & Tucker, 2011). For example, *yesterday* may be realized as [jɛʃeɪ] (Tucker, 2007) and 原因 *genin* 'reason' may be realized as [gẽːin] in Japanese. These examples suggest that the actual pronunciation (phonetic form) of words could be substantially inconsistent with the spelling. This additional inconsistency created by reduced forms is ubiquitous (Dilts, 2013; Johnson, 2004). Of particular interest is the fact that these reduced forms are more difficult to process than canonical/unreduced counterparts despite the fact that reduced forms occur more frequently than canonical/unreduced ones (e.g., Arai, Warner & Greenberg, 2007; Ernestus, Baayen & Schreuder, 2002; Tucker, 2007, 2011; van de Ven, Tucker & Ernestus, 2011). Importantly, recent research has argued that the reason that unreduced forms are easier to process than reduced counterparts could be due to the orthography, particularly the consistent relationship between the unreduced pronunciation and its orthographic form (Bürki, Spinelli & Gaskell, 2012; Charoy & Samuel, 2019; Racine, Bürki & Spinelli, 2014; Ranbom & Connine, 2007, 2011; Viebahn, McQueen, Ernestus, Frauenfelder & Bürki, 2018), suggesting that phonological and orthographic consistency might play an important role in the processing of reduced forms. Mitterer & Reinisch (2015) extended the investigation of the orthographic effect on reduced forms by looking at conversational speech. In their study, reduced forms were presented with conversational-speech like context. They found that orthography does not play an important role in the recognition of conversational speech.

Similar effects of orthography were found in second language (L2) speakers.

Veivo & Järvikivi (2013) and Veivo, Järvikivi, Porretta & Hyönä (2016) demonstrated that both L1 and L2 orthography plays a role in the perception of spoken words in an L2, but the effect of orthography depends on the proficiency of L2 speakers because early learners have not yet fully established phonological and orthographic representations of L2 words.

Logographic languages, such as Chinese and Japanese, also show the effect of phonological and orthographic consistency. In such languages however, the pronunciation and spelling mappings are not straightforward because logographs do not reliably correspond to any sub-lexical phonological unit (Fushimi, Ijuin, Patterson & Tatsumi, 1999; Wydell, 1998; Wydell, Patterson & Humphreys, 1993). While logographic scripts cannot be decomposed into smaller orthographic elements that correspond to sub-lexical phonological units, they can be decomposed into smaller subcharacter components, one of which is the phonetic radical that often provides a clue to the pronunciation of the whole character. The reliability of the phonetic radicals has been utilized to define the pronunciation and orthographic consistency in Chinese. As in alphabetic languages, similar orthographic consistency effects were found in both L1 (Chen, Chao, Chang, Hsu & Lee, 2016; Qu & Damian, 2017) and L2 Chinese speakers (Qu, Cui & Damian, 2018). Research on the P-O consistency effect in Japanese is very limited. Hino, Kusunose & Lupker (2017) investigated the P-O consistency effect using the P-O consistency index which is calculated based on the frequency of phonological and orthographic neighbours of the target words. Hino et al. (2017) found that response latencies for low consistency words were slower than for high consistency words. Similar orthographic effects were also found in L2 Japanese speakers. As in alphabetic languages, the effect was dependent of L2 proficiency, where compared to lower proficiency speakers, higher proficiency speakers looked to the target words more than to the competitors (Mitsugi, 2018).

1.2 Aims of dissertation

This dissertation describes three studies that investigate phonetic variability of speech and how such variability interacts with P-O consistency in spoken word recognition in L1 and L2 Japanese. These studies are presented as separate, self-contained journal-style papers. Previous research has suggested that there are effects of orthography in spoken word recognition in both alphabetic and logographic languages for both L1 and L2 speakers (e.g., Chen et al., 2016; Qu et al., 2018; Veivo et al., 2016; Ziegler & Ferrand, 1998). For L1 speakers in alphabetic languages, the orthographic effects are likely to disappear in conversational speech (Mitterer & Reinisch, 2015). In this dissertation, we further investigate the orthographic effect by focusing on the effect of P-O consistency in Japanese speakers. In particular, we compare (1) the distribution and degree of phonetic reduction between casual and careful speech, and then we examine (2) how orthography affects the processing of casual speech compared to careful speech, (3) how orthography affects processing in a logographic language, and (4) how orthography affects processing in L2 learners of Japanese.

1.3 Outline of the dissertation

This dissertation is structured as follows. Chapter 2 investigates the distribution and degree of phonetic reduction across various styles of speech. Chapters 3 and 4 investigate how the P-O consistency effect interacts with reduced word forms in L1 Japanese (Chapter 3) and how the consistency effect interacts with both reduced word forms and proficiency levels in L2 speakers (Chapter 4). In Chapter 5, I conclude this dissertation by summarising the results of the individual studies and addressing the research aims and questions raised in this dissertation, followed by the limitations and future directions of the studies. A short summary of each study is provided below.

1.3.1 Specific studies

Study 1

We investigated the distribution and degree of phonetic reduction across various styles of speech using the Corpus of Spontaneous Japanese (Maekawa, 2003). Specifically, we examined the effect of speech style on the realization of word-medial stops and word-medial nasals and discuss the acoustic results of phonetic reduction in comparison to English (Warner & Tucker, 2011).

We used the core portion of the Corpus of Spontaneous Japanese (CSJ). The CSJ contains 662 hours of speech and about 7.5 million words (Maekawa, 2003). The core portion is a subset of the CSJ, including about a half million words corresponding to about 44 hours of speech in four different styles of speech: Academic Presentation, Simulated Public Speech, Dialogue, and Read Speech. Academic Presentation includes live recordings of academic talks. Simulated Public Speech includes studio recordings on everyday topics, presented in front of a small audience in a relatively relaxed atmosphere. Dialogue is composed of interviews, task oriented dialogue, and free dialogue. Transcriptions of the Academic Presentation were read by the same speaker for Read Speech. We used the phonological transcriptions provided in the dataset. Using Praat (Boersma & Weenink, 2016), we extracted all intervals that contained the target segments, word-medial voiced stops and word-medial nasals.

Using logistic and linear mixed-effects modelling, we found a gradual effect of reduction across speech styles particularly when using intensity difference as a measure of reduction, where the most spontaneous speech demonstrates the greatest reduction, and the least spontaneous speech exhibits the smallest reduction for both voiced stops and nasals. However, we did not find a clear gradual effect as Maekawa (2005) suggested. This could be because the effect of speech style is not as strong as expected especially when other confounding factors are held constant.

For voiced stops, when we used the presence or absence of complete closure as a measure of reduction, we found an effect of speech style, phoneme, and their interaction, indicating less reduction for the least spontaneous speech than for

the most spontaneous speech. This result is in line with what has been found in Warner & Tucker (2011) demonstrating greater reduction for more a spontaneous speech style in English. Using duration as a parameter, an effect of speech style, phoneme, and their interaction also appeared, which is in line with the findings of Warner & Tucker (2011). However, the direction of the effect differs. Whereas Warner & Tucker (2011) found greater reduction for more spontaneous speech, we found the least reduction in the most spontaneous speech. This could be due to the fact that our Japanese corpus comprises speech styles that lean towards fast speech, leading to shorter segment duration. Using intensity difference as a measurement, we found the most consistent effects of speech style. Intensity difference results followed the predicted reduction pattern, where the most spontaneous speech demonstrates the greatest reduction. However, the pairwise comparisons between speech styles reached significance only between the smallest and largest value of intensity difference. The result of our study is in agreement with Warner & Tucker (2011) and Maekawa (2005), in that we all demonstrate that more spontaneous speech contains greater reduction.

With duration as a measure of reduction of nasals, while the overall effect of speech style and the interaction between speech style and phoneme resulted solely from /N/, all phonemes demonstrated a tendency of the predicted reduction pattern, the least spontaneous speech indicating the smallest reduction. Similar to the results of voiced stops, the effect of speech style was better reflected in the intensity difference for nasals. While further analysis of /N/ considering the different assimilation environments is necessary to better understand reduction for this segment, this result extends the findings of Warner & Tucker (2011) in a way that the role of speech style on the realization of nasals is reflected in a similar way to that of stops and flaps.

Overall, the acoustic results of reduction and their distributional patterns across speech styles for voiced stops are comparable between Japanese and English, and nasals also indicate comparable reduction patterns. These results extend the findings of Barry & Andreeva (2001) that there are language-independent patterns of reduction regardless of language.

Study 2

On the basis of the results of Study 1, we created stimuli to examine how the effect of P-O consistency interacts with reduced word forms in Japanese. Specifically, we compared the time course of the consistency effect between reduced and unreduced word forms in a Go-NoGo and delayed naming task. In our Go-NoGo task, participants responded to non-target items by pressing a button on a pad, and they did not respond to the target items. In this way, the target trials were free of artifacts resulting from motor movements invoked by responses that possibly influence pupil dilation, and the participants did not need to make any linguistically derived decisions on the the target items. In our delayed naming task, participants heard a Japanese word and waited until they heard a pure tone. They then repeated what they had heard.

In the present study, we conducted pupillometry experiments. Pupil dilation represents a physiological correlate of cognitive load, and pupillometry, the measurement of pupil dilation, has been utilized as an index of cognitive effort (Kahneman & Beatty, 1966). Pupillometry offers a reliable method to measure cognitive effort, attention, and affect imposed by different variables in speech processing in the absence of voluntary processes (Laeng, Sirois & Gredebäck, 2012; Papesh & Goldinger, 2012, 2015) and it is also largely free from the effect of task-specific strategies (Goldinger & Papesh, 2012). The locus coeruleus activities are correlated with two modes of pupillary responses: phasic and tonic. While the phasic pupillary response is time locked to task-related events and stimuli, the tonic (or baseline) response is slow-changing and is related to the state of arousal or vigilance (Laeng et al., 2012; Papesh & Goldinger, 2015).

We used 226 four-mora and two-logograph Japanese words. Following the results of corpus analysis in Chapter 2, all target words contained a word-medial nasal and/or voiced stop and were recorded in both reduced and unreduced forms by a female native Japanese speaker (452 items in total). We instructed the speaker to produce the words clearly for unreduced forms and casually (spontaneous-speech like) for reduced forms. The speaker produced multiple tokens of both

forms, and we selected the most natural sounding tokens as stimuli. Overall, reduced forms have shorter duration, faster speech rate, lower mean pitch, and a smaller intensity difference than unreduced counterparts. The differences between the two forms reached statistical significance for all the properties except the intensity difference.

We used the Balanced Corpus of Contemporary Written Japanese (Maekawa, Yamazaki, Ogiso, Maruyama, Ogura, Kashino, Koiso, Yamaguchi, Tanaka & Den, 2014) to obtain information needed to calculate the P-O consistency index for each target word, including word frequency and the number of phonological and orthographic neighbours. Phonological neighbours were defined as words that differ by a single mora from the target word, and orthographic neighbours were defined as words that differ by a single character from the target word (Fushimi et al., 1999).

We applied Generalized Additive Mixed Modeling (GAMM) to our pupillometic data for three reasons (Hastie & Tibshirani, 1990; Wood, 2006). First, GAMM allows us to model non-linear relationships between a dependent and independent variables (Sóskuthy, 2017; Wieling, Tomaschek, Arnold, Tiede, Bröker, Thiele, Wood & Baayen, 2016). This was important as we expected pupil size to fluctuate over time (van Rij, Hendriks, van Rijn, Baayen & Wood, 2019). Second, GAMM can model multiple dimensional nonlinear interactions of continuous variables. Third, GAMM allowed us to control for serial dependency in time-series data, namely, autocorrelation (Baayen, Vasishth, Bates & Kliegl, 2017).

We found that (1) reduced forms elicited larger dilation than unreduced counterparts, (2) the P-O consistency effect emerged in both tasks, and it influences reduced and unreduced forms differently, (3) dilation becomes greater as the P-O consistency increases. Reduced word forms elicited greater pupil dilation than unreduced word forms in both tasks, and for the Go-NoGo task, while we found the effect of P-O consistency for both forms, the direction of the effect differed between the two forms, suggesting that the consistency effect influenced the two forms differently. Our results also suggest that regardless of the explicitness of tasks, orthography plays a role for the comprehension of careful speech (unreduced form). In contrast to the results of Mitterer & Reinisch (2015), we found

the effect of orthography in an implicit task with casual speech (reduced forms). This could be because of the difference in the degree of "conversational-likeness". While reduced forms are presented with informal sentences including discourse markers and contractions in their study, these forms are presented in isolation in our study. This result suggests that while the orthography plays a role in the recognition of reduced forms in isolation, the effect disappears when such forms are presented with more conversation-like context, suggesting that the way in which reduced forms are presented plays a more important role than the explicitness of tasks. For the delayed naming task, while we observed the P-O consistency effect for both forms, the effect was greater for reduced forms. This suggests that the P-O consistency effect influenced the two forms differently and that the consistency effect played a more important role when requiring more processing cost (reduced forms).

The direction of the consistency effect found in unreduced forms was opposite to what has been found in Hino et al. (2017) and other previous studies (e.g., Ziegler & Ferrand, 1998; Ziegler et al., 2004). This could be caused by correlations between the consistency effect and other lexical variables. Further analysis of our results have revealed that low P-O consistency words tend to have a higher number of phonological neighbours. Yoneyama (2002) shows that high phonological neighbourhood density facilitates the recognition of Japanese words, meaning that the facilitatory effect of high phonological neighbourhood density competes with the inhibitory effect of low P-O consistency. Another reason for the opposite direction of the consistency effect could be the fact that the direction of P-O consistency effect changes depending on the type of orthography. Previous research has shown that in French (alphabetic), inconsistent words elicited increased negativity (N400) (Pattamadilok et al., 2009; Perre et al., 2011), but in Chinese (logographic), the effect is reversed such that consistent words elicited increased negativity (Chen et al., 2016).

Study 3

We extended the investigation of Study 2 and examine how the P-O consistency effect interacts with phonetic reduction for L2 Japanese speakers of L1 English. Following Study 2, we employed the same stimuli (452 items), measure (pupillometry), tasks (Go-NoGo and delayed naming), and statistical approach (GAMM) in order to compare the time course of the P-O consistency effect between reduced and unreduced Japanese words. We found that (1) reduced forms elicited larger dilation than unreduced counterparts, (2) the P-O consistency effect emerged in both tasks, and it influences reduced and unreduced forms differently, (3) the consistency effect also varies depending on L2 proficiency, and finally (4) dilation becomes greater as the P-O consistency increases.

First, as in previous studies in L1 (e.g., Ernestus et al., 2002; Tucker, 2011), we observed that unreduced forms are easier to process than their reduced counterparts in both tasks. However, the difference between the two forms was not as large as expected. This could be due to P-O consistency. We found an overall trend that the difference between the two forms is greater for consistent words than for inconsistent words. These results suggest that P-O consistency plays an important role in the recognition difficulty of reduced word forms. This is in line with the discussion of orthographic effects in the recognition of reduced variants in previous studies (Bürki et al., 2012; Charoy & Samuel, 2019; Racine et al., 2014; Ranbom & Connine, 2007, 2011; Viebahn et al., 2018).

Second, we observed the effect of P-O consistency for both tasks although previous studies demonstrated that the effect was null in a naming task (Ventura, Morais & Kolinsky, 2007; Ventura, Morais, Pattamadilok & Kolinsky, 2004; Ziegler et al., 2004). This difference could be due to the measurement utilized. Whereas earlier studies on the P-O consistency effect employed offline measures (e.g., reaction latency, accuracy rate), later research tends to use online measures (e.g., ERPs, sLORETA), and the studies that employed a naming task utilized the offline measure, naming latency. Our application of pupillometry to a naming task has revealed that the consistency effect emerges in the task despite the fact that the

effect was modest.

Third, our results demonstrate an interaction between reduction, L2 proficiency, and P-O consistency over time. However, the interaction pattern found in the Go-NoGo task displays a different direction compared to what has been found in Veivo et al. (2016). This could be due to the difference in tasks, where orthographic information was visually presented to participants in Veivo et al. (2016) but it was not in our study.

Fourth, similar to the results of Chapter 3, the direction of the consistency effect found in our results was opposite to what has been found in Hino et al. (2017) and other previous studies (e.g., Ziegler & Ferrand, 1998; Ziegler et al., 2004). The same discussion in Chapter 3 above applies here as well.

Lastly, our study represents the applicability of pupillometry to examine such effects in spoken word recognition.

1.3.2 Summary

All studies together, we investigate how production of speech varies across different styles of speech and how listeners interact with such variability, and how visual language information interacts with auditory language information during spoken language comprehension. The results of these studies will contribute to our understanding of how humans process spoken language using multiple sources of language information to cope with variability.

In the section follows, Chapter 2 describes Study 1 investigating the distribution and degree of phonetic reduction across various styles of speech. Chapters 3 and 4 describe Study 2 investigating how the P-O consistency effect interacts with reduced word forms in L1 Japanese and Study 3 examining how the consistency effect interacts with both reduced word forms and proficiency levels in L2 speakers.

Chapter 2

Phonetic variability of voiced stops and nasals in Japanese across various styles of speech

2.1 Introduction

The realization of segments in words is highly variable (e.g., Ernestus & Warner, 2011). Previous research has found that more than 60% of words in the Corpus of Conversational American English (Pitt, Dilley, Johnson, Kiesling, Raymond, Hume & Fosler-Lussier, 2007) were realized in variable forms, and 25% of these forms demonstrated segment deletion as compared to a dictionary transcription of the word (Dilts, 2013; Johnson, 2004). This variability is often the result of phonetic reduction, realized as shortening, deletion, and/or incomplete articulation of segments (Greenberg, 1999; Warner & Tucker, 2011). For example, *yesterday* may be realized as [jɛʃeɪ] (Tucker, 2007) and *but I was like* may be as realized as [bˌɹʌʒləɪ̃] (Warner, 2011). In Japanese, *sukoshi hanashite* 'speak a little bit' may be realized as [sɯkoʃɑnɑʃite] (Arai, 1999). While such reduction is often considered to be governed by constraints of both articulation and perception (Lindblom, 1990), research has shown that the distribution, degree and type of reduction may vary depending on language and/or style of speech (Ernestus & Warner, 2011; Keune, Ernestus, van Hout & Baayen, 2005; Torreira & Ernestus, 2011; Warner & Tucker, 2011). In the present study, we compare the distribution and degree of phonetic reduction across various styles of speech in Japanese. More specifically, we investigate the

effect of speech style on the realization of word-medial stops and word-medial nasals and discuss the acoustic results of reduction in comparison to Warner & Tucker (2011) who have investigated the effect of speech style on the realization of stops (and flaps) in English.

When the terms "spontaneous/casual speech" and "read speech" are used, these terms often refer to polarized styles of speech. Spontaneous/causal speech generally stands for informal, dynamic, and unrehearsed speech as in a casual conversation while read speech refers to scripted, careful, and formal speech as in a student reading a list of words (Laan, 1997; Labov, 1972). However, if we consider a politician spontaneously giving a speech or an actor reading a script, it is clear that their speech will often deviate from the typical characteristics of spontaneous and read speech (Haynes, White & Mattys, 2015). In order to illustrate the dynamics of speech style, speech style can be considered a continuum ranging from careful to casual speech (Tucker, 2007). Careful speech, often described as hyper-articulated speech, holds at one end, and such speech may be directed toward non-native and/or the hard of hearing listeners. The other end of the continuum is held by casual speech, also referred to as hypo-articulated, spontaneous, or conversational speech (Lindblom, 1990; Warner, 2012). In comparison to careful speech, casual speech exhibits a higher articulation rate, lower F0 variation and F0 declination, more frequent hesitations, approximated articulation, and shorter segment duration and prosodic units (Cutler, 1998; Laan, 1997; Mehta & Cutler, 1988).

Such characteristics of casual speech are often connected to phonetic reduction, and research has shown that there is a relationship between the spontaneity of speech and the distribution and/or degree of reduction (Warner & Tucker, 2011). Maekawa (2005) found in the Corpus of Spontaneous Japanese (CSJ) (Maekawa, 2003) that 20% of type tokens include pronunciation variants, and the occurrence ratio of these variants differs depending on a speech style; the occurrence ratio of pronunciation variants is highest in dialogue, second highest in simulated spontaneous speech, third highest in academic presentation, and lowest in read speech. The researcher argues that this order follows the degree to which speakers are

conscious about the way they speak, suggesting that dialogue is the most casual speech and that read speech is the most careful speech. Using Mel Frequency Cepstrum Coefficients, Nakamura, Iwano & Furui (2008) compared the spectral characteristics of spontaneous speech to read speech using two corpora: CSJ and JNAS, the Japanese Newspaper Article Sentences (Itou, Yamamoto, Takeda, Takezawa, Matsuoka, Kobayashi, Shikano & Itahashi, 1999). Results revealed that there is a systematic correlation between speech spontaneity and the shrinkage of spectral space; that is, spectral space shrinks as the speaking style becomes more spontaneous (i.e., more casual).

While studies have shown that phonetic reduction occurs in similar ways cross-linguistically as in the manifestation of shorter segment duration, weakened or deleted segments, centralized vowels, shrunk spectral space (Barry & Andreeva, 2001; Laan, 1997; Nakamura et al., 2008; Van Son & Pols, 1999; Warner & Tucker, 2011), other studies have shown that reduction patterns may differ between languages or dialects (Keune et al., 2005; Torreira & Ernestus, 2011) and that the effect of correlation between the spontaneity of speech and the distribution and/or degree of reduction may differ (Keune et al., 2005; van Dommelen, 2018). In the present study, by comparing the distribution and degree of phonetic reduction between speech styles in Japanese, we also illustrate how comparable the acoustic results of reduction are between Japanese and English.

Our focus is on the phonetic reduction of word-medial voiced stops and word-medial nasals as it has been reported that these types of segments have been phonetically reduced in various forms in Japanese (Arai, 1999; Arai et al., 2007; Maekawa, 2003; Vance, 2008). For example, in the dataset of Japanese spontaneous speech in the OGI Multi-Language Telephone Speech Corpus (Muthusamy, Cole & Oshika, 1992), the word /tenisu/ 'tennis' was realized as [tẽːsɯ]; the word-medial nasal /n/ was completely deleted, and the following vowel /i/ was also deleted (Arai, 1999). Additionally, the preceding vowel /e/ was both nasalized and lengthened (Arai, 1999). For voiced stops, Arai (1999) found the instances where articulation of the word-medial voiced stop /ɡ/ in /daigakɯ/ 'university' was approximated due to the lack of full oral closure and realized as [daiɣakɯ], and in extreme cases,

the consonant was completely deleted and realized as [dɑiɑkɯ]. Furthermore, Arai (1999) found that among all occurrences of phoneme /g/ in the dataset, 20.4% of them appeared with a clear burst, 72.2% of them were realized as [ɣ] without a clear burst, and 7.4 % of them alternated to [ŋ].

In the present study, we use a large-scale speech corpus, the Corpus of Spontaneous Japanese (Maekawa, 2003), and categorize phonetic reduction into two types, "reduction" and "deletion". We view reduction as approximation or modification of articulatory features, resulting in variable acoustic realizations of segments, and we treat deletion as the lack of the realization of acoustic features, resulting in deletion of segments. We focus on the former type, reduction, because of the following reasons. First, we are interested in the extent to which acoustic features vary due to the reduction. In the case of deletion, there is no acoustic feature to measure. Second, there appears to be a categorical difference between the processes of phonetic reduction and segment deletion. The result of phonetic reduction processes is often measured by the degree to which acoustic features of the target segment vary, and segment deletion is viewed as an extreme case of phonetic reduction. However, Turnbull (2018) argues that the difference between phonetic reduction and segment deletion may be more categorical because the environment where segment deletion occurs does not necessarily result from reduction. We therefore believe that it is reasonable to examine these two types of reduction processes separately.

We analyzed the duration and intensity difference of target segments across four styles of speech: Academic Presentation, Simulated Public Speech, Dialogue, and Read Speech. The intensity difference was defined as the difference between the minimum intensity of the target segment to the averaged maximum intensity of surrounding segments (Tucker, 2011; Warner & Tucker, 2011). On the basis of the relationship between the spontaneity of speech and the degree of phonetic reduction revealed by previous studies, we expect stronger reduction (more approximant-like productions), as indicated by shorter durations and smaller intensity differences, as speech style becomes more spontaneous. Following the order of degree to which speakers are conscious about the way they speak (Maekawa,

2005), the shortest duration and the smallest intensity difference should be found in Dialogue (most spontaneous), which should be followed by Simulated Public Speech. The third shortest duration and the third smallest intensity difference should be observed in Academic Presentation, and the longest duration and the largest intensity difference should be found in Read Speech (least spontaneous).

2.2 Method

2.2.1 Corpus data

We used the core portion of the Corpus of Spontaneous Japanese, which comprises four styles of speech: Academic Presentation, Simulated Public Speech, Dialogue, and Read Speech. Academic Presentation includes live recordings of academic talks. Simulated Public Speech includes studio recordings on everyday topics, presented in front of a small audience in a relatively relaxed atmosphere. Dialogue is composed of interviews, task oriented dialogue, and free dialogue. Transcriptions of Academic Presentation were read by the same speaker for Read Speech. An overview of the dataset is provided in Table 2.1. The size of the Dialogue and Read Speech data is smaller than that of Academic Presentation and Simulated Public Speech data.

Table 2.1: An overview of the corpus: Styles (Dialogue; Simulated Public Speech; Academic Presentation; Read Speech); Talks (Number of talks); Speakers (Number of speakers); Hours (Total hours of talks); Vstop (Number of sampled voiced stops); Nasal (Number of sampled nasals); Total Samples (Total number of sampled voiced stops and nasals.

Styles	Talks	Speakers	Hours	Vstops	Nasals	Total Samples
Academic Presentation	70	58	14.2	20498	14996	35494
Simulated Public Speech	107	75	15.0	21939	14528	36467
Dialogue	18	6	3.0	1176	1044	2220
Read Speech	6	6	1.4	1084	815	1899

We used the phonological transcriptions provided in the dataset. These transcriptions are generated and time-aligned using an HMM-based alignment technique and adjusted by human labellers. In the dataset, voiced stops are labelled as **/<cl>/** for closure intervals and as **/b**, **d** or **g/** for the burst release intervals that follows. The closure intervals are defined from the offset of a previous segment to the

offset of the closure interval. We added the duration of the closure and subsequent burst release intervals together to measure the duration of entire stops. Frequently, these word-medial stops are realized without a complete closure and therefore no clear burst release occurs. In such cases, these closure and burst release intervals are labelled together as a single unit as **/<cl>,b/**, **/<cl>,d/** or **/<cl>,g/**. Furthermore, phonologically palatalized voiced stops, transcribed as **/by/**, were excluded from our analysis because these are separate phonemes. In contrast, phonetically palatalized voiced stops, transcribed as **/bj/**, were collapsed and treated as **/b/** because these are allophonic variations of the voiced stop consonants (Kawahara, 2017). For nasals **/m, n and N/**, we applied the same procedures for the palatalization for **/m/ and /n/**.

2.2.2 Procedures and materials

Using Praat (Boersma & Weenink, 2016), we extracted all intervals that contained the target segments. Some of the intervals consisted of multiple labels, suggesting that these segments were reduced. For example, Figure 1 depicts a waveform and spectrogram of *dankai desu node*, 'Because [it] is a stage where ...', realized as *[daNkaisnode]*. We can observe that *desu* is merged into a single interval due to the reduction of /d/ (no complete closure or burst release), /e/ (almost deleted), and the final /u/ (entirely deleted). As seen in the alignment of these segment labels in the figure, the deleted or reduced segments were labelled together with neighboring ones. We excluded these highly reduced segments from our analysis because it is impossible to obtain accurate measures of the acoustic properties of target segments. We also excluded highly reduced pronunciation variants, where target segments were realized as different phonemes. For example, 31% of *ni* in *nichi*, 'day' was transcribed as *N* as in *Nchi*. We found 416 instances of such reduction, and many of them were function words or high frequency words (Bybee, 2007). We also found three lexical items, whose tokens are reduced more than 70% of the time: **geNiN** transcribed as **geHiN** (94%), **zeNiN** transcribed as **zeHiN** (81%), and **teNiN** transcribed as **teHiN** (75%). The reduction process for these words seems to be lexicalized (Bybee, 2007).

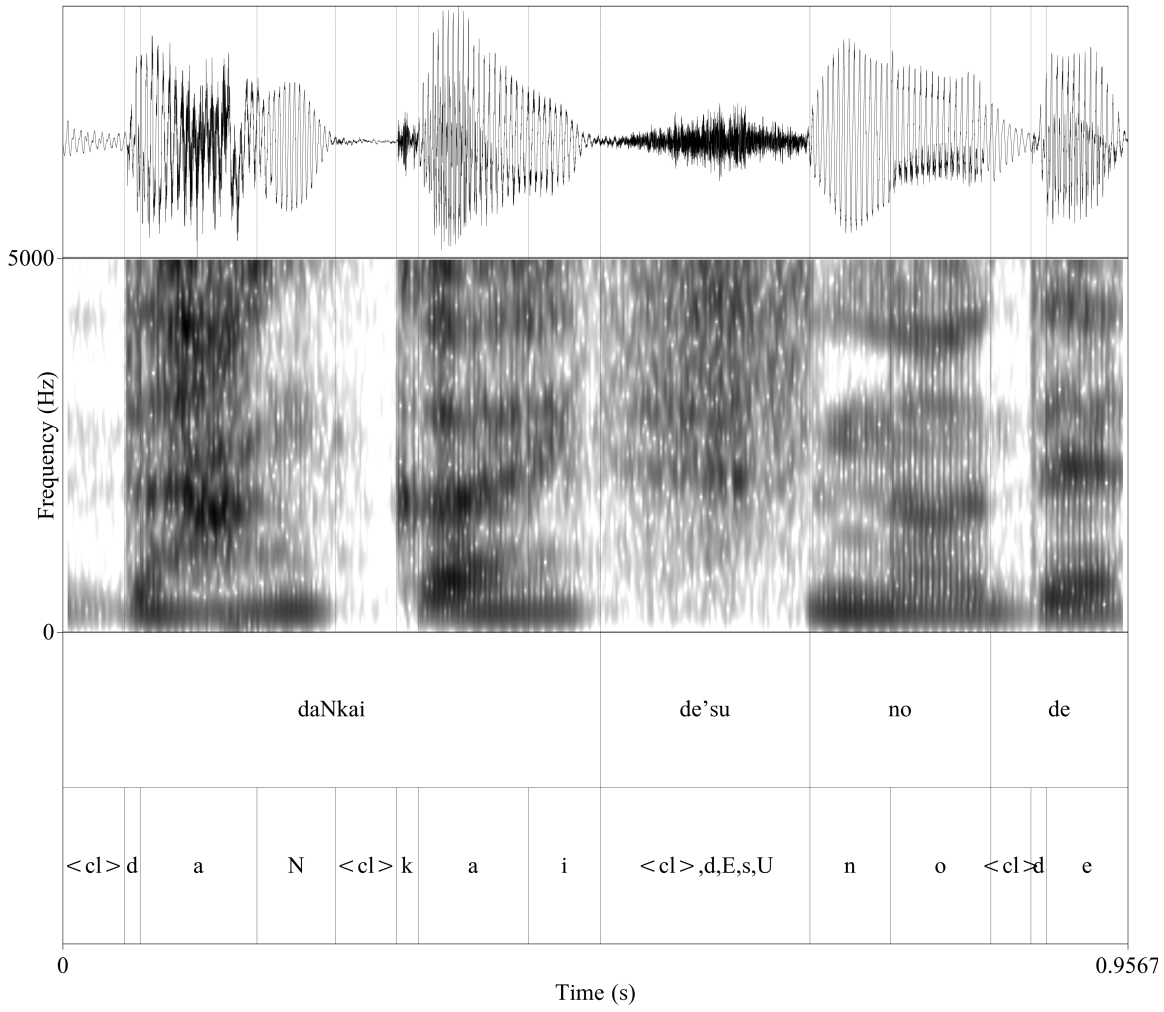

*Figure 2.1: Waveform and spectrogram of **dankai desu node**, "Because [it] is a stage where . . . ", realized as **[daNkaisnode]**.*

After visually inspecting the range of segment duration for each phoneme, we excluded segments with a duration that was more than 2.5 standard deviations apart from the mean duration for each phoneme. Finally, we also excluded target segments that occurred in extremely low frequency words. If the word was not listed in the Balanced Corpus of Contemporary Written Japanese (BCCSJ) (Maekawa et al., 2014), we excluded it. We chose the BCCSJ for this process because it is one of the largest Japanese corpora, and it encompasses a wide range of styles and genres. As shown by Table 2.1, the number of sampled target segments for Dialogue and Read Speech is smaller than that of Academic Presentation and Simulated Public Speech. In order to deal with these unbalanced datasets, we randomly selected 2500 samples from the Academic Presentation and Simulated Public Speech datasets (approximately 10% of the original data), and used these samples for our analysis. We conducted our analyses with both original and randomly sampled datasets and found similar effects. In the present study, we only report results of randomly sampled datasets.

2.3 Results

In the sections that follows, we conducted logistic and linear mixed-effects modelling in R (R Development Core Team, 2018) using the R package LME4 (Bates, Mächler, Bolker & Walker, 2015) with p-values computed with Satterthwaite approximations to degrees of freedom with LMERTEST package (Kuznetsova, Brockhoff & Christensen, 2017). We also used EMMEANS package to report the estimated marginal means for the variables of interest (Lenth, 2019). Three response variables (the presence or absence of a complete closure, duration, and intensity difference of target segments) were fitted separately as a function of two predictor variables (speech style and phoneme) with control variables that could affect these response variables. We employed a backwards stepwise elimination procedure for fixed effects (Matuschek, Kliegl, Vasishth, Baayen & Bates, 2017). For random effects, we included word and speaker identity as random intercepts (Baayen, Davidson & Bates, 2008). We discuss the statistical analysis of each response variable, along

with the descriptions of control variables, as follows: first, the presence or absence of a complete closure, duration, and intensity difference of voiced stops, and second, duration and intensity difference of nasals.

2.3.1 Presence or absence of a complete closure in voiced stops

In the CSJ dataset, a closure duration of word-medial voiced stops is segmented from the offset of a previous segment to the onset of the subsequent burst release. Sometimes, these stops are realized without a complete closure. In such cases, a closure and burst release were labelled together as a single unit (i.e., reduced forms). Overall, 58.94% of word-medial voiced stops were realized without a complete closure. The number of occurrences of /d/ was highest (1606 times), but the percentage that /d/ was realized without a complete closure was lowest (34.56%). The number of occurrences of /g/ was lower than that of /d/ (1395 times), but the percentage that /g/ was realized without a complete closure was highest for /g/ (83.30%). The number of occurrences of /b/ was lowest (842 times), and its percentage that /b/ was realized without a complete closure was lower than /g/ but higher than /d/ (65.08%). As for speech styles, the percentage of the stops without a complete closure was highest in Academic Presentation (60.91%), second highest in Dialogue (60.57%), third highest in Simulated Public Speech (57.57%), and lowest in Read Speech (56.32%). The percentage of the stops without a complete closure between phonemes appears to vary more than between speech styles.

In the following analysis, we used logistic mixed-effects modelling to compare the likelihood of the presence or absence of a complete closure for voiced stops across the phonemes and speech styles. Our final models included word and speaker identity as random intercepts, and speech rate as a control variable. Speech rate demonstrated that the likelihood of the absence of a complete closure increased as speech became faster. Pairwise comparisons across the phonemes indicated that the likelihood of the presence or absence of a complete closure was different among the phonemes. That is, the likelihood of the absence of a complete closure of /g/ was higher than that of /b/ (z=7.891, $p<.0001$) or that of /d/ (z=16.825, $p<.0001$), and the likelihood of the absence of a complete closure of /b/

was higher than that of /d/ (z=9.728, $p<.0001$). However, none of the pairwise comparisons across speech styles reached significance, suggesting that the type of phonemes plays a more important role than the style of speech in the likelihood of presence or absence of complete closure.

We also investigated the interaction effect between the phonemes and speech styles, revealing that including the interaction improved the model (χ^2=24.515, $p<.001$). In order to further examine the interaction effect, we divided the data into separate phonemes and modeled them individually. Our final model for each phoneme included the same random intercepts and control variables as the overall model. Table 2.2 illustrates the overview of the percentage of absence of a complete closure for each phoneme and speech style. For /b/, the mean percentage of absence was highest in Dialogue, second highest in Academic Presentation, third highest in Simulated Public Speech, and lowest in Read Speech. Pairwise comparisons among the speech styles revealed that the comparison between Dialogue (highest ratio) and Read Speech (lowest ratio) reached significance (z=2.986, $p<.005$) for /b/, meaning that the likelihood of absence of a complete closure was higher in Dialogue than in Read Speech. The mean percentage of absence of /d/ was highest in Academic Presentation, second highest in Simulated Public Speech,, third highest in Dialogue, and lowest in Read Speech. While the percentage of absence of /b/ was highest in Dialogue, the percentage of /d/ was second lowest in Dialogue. This different pattern of percentage of absence could be due to the fact that the overall percentage of absence in /d/ was much lower than in /b/. Pairwise comparisons among the speech styles indicated that none of the comparisons reached significance for /d/. As for /g/, the mean percentage of absence was highest in Simulated Public Speech, second highest in Dialogue, third highest in Academic Presentation, and lowest in Read Speech. More than 80% of /g/ lacked a complete closure in all styles. Pairwise comparisons among the speech styles revealed that the comparison between Simulated Public Speech and Academic Presentation reached significance (z=2.767, $p<.05$). As a result, we found that the least spontaneous speech, Read Speech, demonstrates the lowest absence ratio for all phonemes (Table 2.2).

Table 2.2: Overview of occurrence of voiced stops and the percentage of absence of a complete closure for phonemes and speech styles. Segment (target segments); Speech Style (Dialogue; Simulated Public Speech; Academic Presentation; Read Speech); Total Occurrence (total occurrence of target segment); Absence% (The percentage of the absence of complete closure); Rank (1 indicates the highest and 4 indicates the lowest absence percentage).

Segment	Speech Style	Total Occurrence	Absence%	Rank
b	Dialogue	265	70.94	1
b	Simulated Public Speech	262	61.07	3
b	Academic Presentation	183	70.49	2
b	Read Speech	132	53.79	4
d	Dialogue	407	33.42	3
d	Simulated Public Speech	460	34.35	2
d	Academic Presentation	404	37.87	1
d	Read Speech	335	32.24	4
ɡ	Dialogue	345	84.64	2
ɡ	Simulated Public Speech	322	87.89	1
ɡ	Academic Presentation	380	80.79	3
ɡ	Read Speech	348	80.46	4

The absence of a complete closure as an index of reduction is binary, but a reduction process can also be gradient indicated by a shorter duration or smaller intensity difference. Although /d/ did not display the binary reduction in as many instances as the other stops, it might demonstrate gradient reduction processes. The following section therefore investigates the duration and intensity difference of word-medial voiced stops across speech styles.

2.3.2 Duration of voiced stops

In the following analysis, we used linear mixed-effects modelling to compare the duration of voiced stops across phonemes and speech styles, using all samples (with and without a complete closure). We included word and speaker identity as random intercepts and several control variables that could affect the segment duration. These include the duration of words containing the segments, the frequency of words containing them, the type of mora containing them (accented or not), and the rate of speech. Our final models comprised the two random intercepts, with the duration of words and speech rate as control variables, indicating shorter word duration and faster speech rate for shorter segment duration. The mean segment duration was longest in /b/, second longest in /d/, and shortest in /ɡ/. Pair-

wise comparisons between the phonemes revealed that the duration of /b/ was longer than that of /d/ (t=7.791, $p<.0001$) or that of /g/ (t=8.893, $p<.0001$), but the comparison between /d/ and /g/ did not reach significance. For speech styles, the mean segment duration was longest in Dialogue, second longest in Simulated Public Speech, third longest in Read Speech, and shortest in Academic Presentation. Pairwise comparisons across the speech styles indicated that the comparisons between Dialogue and Academic Presentation, between Dialogue and Read Speech, or between Simulated Public Speech and Academic Presentation reached significance. That is, the duration of the segments was longer in Dialogue than in Academic Presentation (t=3.225, $p<.005$), or in Read Speech (t=3.854, $p<.005$). Additionally, the duration of the segments was longer in Simulated Public Speech than in Academic Presentation (t=2.634, $p<.05$).

We also investigated the interaction effect between the phonemes and speech styles, revealing that including the interaction improved the model (χ^2=19.2, $p<.005$). For further analysis, we divided the data into different phonemes and modeled them separately. Our final model for each phoneme included the same random intercepts and control variables except for /d/ excluding the speech rate. Figure 2.2 shows the distribution of duration of voiced stops for each speech style and phoneme. The figure reveals that distribution and range of duration appears to be wider for /b/ and /g/ than for /b/, which is reflective of the high reduction ratio of /b/ and /g/. That is, frequent phonetic reduction leads to a wider range of durational variation.

Table 2.3 illustrates the overview of segment durations for each phoneme and speech style. For /b/, the mean duration was longest in Dialogue, second longest in Read Speech, third longest in Simulated Public Speech, and shortest in Academic Presentation. The pairwise comparisons across the speech styles indicated that none of the comparisons reached significance. The mean duration of /d/ was longest in Simulated Public Speech, second longest in Dialogue, third longest in Read Speech, and shortest in Academic Presentation. Pairwise comparisons among the speech styles revealed that only the comparison between Dialogue and Read Speech reached significance (t=3.405, $p<.005$). The segment duration

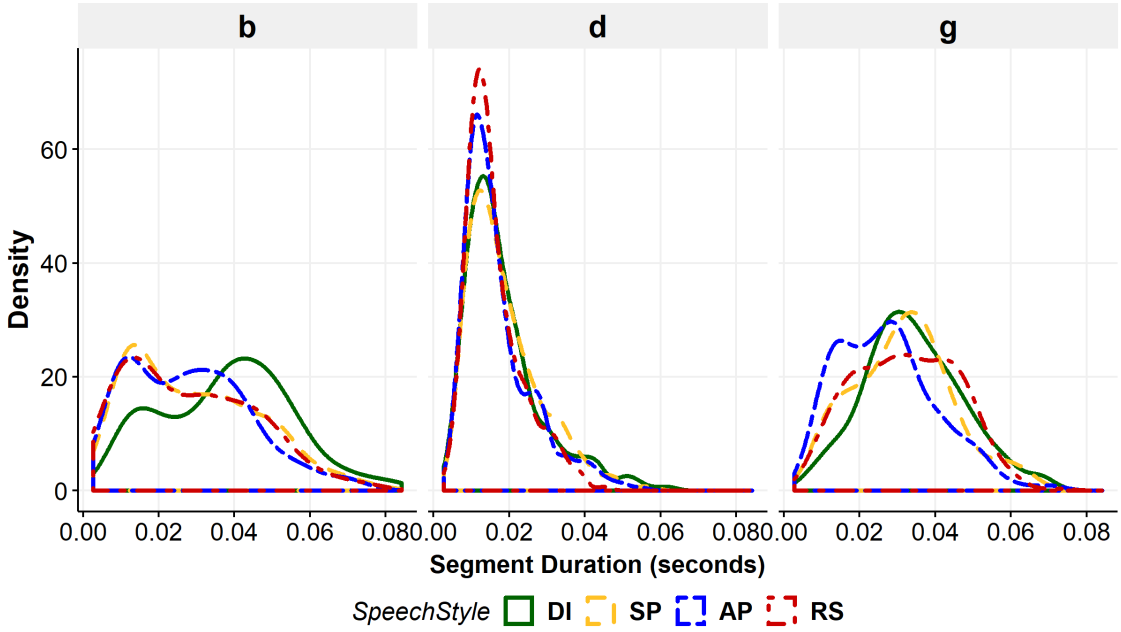

Figure 2.2: The distribution of duration of voiced stops for each phoneme and speech style (DI: Dialogue; SP: Simulated Public Speech; AP: Academic Presentation; RS: Read Speech)

estimated by the model differed from the mean duration, and the estimated segment duration was longest in Dialogue and shortest in Read Speech. As for /g/, the mean duration was longest in Dialogue, second longest in Read Speech, third longest in Simulated Public Speech, and shortest in Academic Presentation. Furthermore, none of the pairwise comparisons across speech styles reached significance.

In sum, we found the shortest mean duration in Academic Presentation and the longest estimated duration in Dialogue for all phonemes. However, since none of the pairwise comparisons across speech styles reached significance for /b/ and /g/ (and only the comparison between Dialogue and Read Speech reached significance for /d/), stylistic differences reflected in the segment duration are small after keeping other factors that could affect the segment duration constant. We also found the extent to which reduction occurs across speech styles differed depending on a measurement of reduction. That is, the reduction pattern found in the binary measure (i.e., presence or absence of a complete closure) was dissimilar to the pattern observed in the temporal measure (i.e., segment duration). While

reduction was greatest (i.e., highest percentage of absence of a complete closure and shortest duration) in Academic Presentation in both measures, reduction was least in Read Speech for the binary measure (i.e., lowest percentage of absence of a complete closure) and in Dialogue for the temporal measure (i.e., longest duration).

Table 2.3: Means and standard deviations (SD) of duration (in seconds) and intensity difference (IntDiff in dB), as well as estimated values of duration and intensity difference by the models, for phonemes and speech styles. For Style, Dialogue; Simulated Public Speech; Academic Presentation; Read Speech. In Rank and Est.Rank columns, 1 indicates the shortest duration and smallest intensity difference and 4 indicates the longest duration and largest intensity difference.

Segment	Style	Parameter	Mean	SD	Rank	Est.Value	Est.Rank
b	Dialogue	Duration	0.0487	0.0152	4	0.0456	4
b	Simulated Public Speech	Duration	0.0446	0.0157	2	0.0440	2
b	Academic Presentation	Duration	0.0392	0.0153	1	0.0410	1
b	Read Speech	Duration	0.0461	0.0152	3	0.0454	3
d	Dialogue	Duration	0.0374	0.0152	3	0.0367	4
d	Simulated Public Speech	Duration	0.0375	0.0139	4	0.0346	3
d	Academic Presentation	Duration	0.0346	0.0144	1	0.0331	1
d	Read Speech	Duration	0.0347	0.0136	2	0.0331	1
g	Dialogue	Duration	0.0379	0.0130	4	0.0366	4
g	Simulated Public Speech	Duration	0.0351	0.0134	2	0.0360	3
g	Academic Presentation	Duration	0.0322	0.0138	1	0.0347	1
g	Read Speech	Duration	0.0378	0.0128	3	0.0352	2
b	Dialogue	IntDiff	10.533	4.806	1	10.467	1
b	Simulated Public Speech	IntDiff	10.834	4.383	2	10.925	2
b	Academic Presentation	IntDiff	10.965	4.218	3	11.304	3
b	Read Speech	IntDiff	11.181	4.690	4	11.856	4
d	Dialogue	IntDiff	10.450	5.683	1	9.975	1
d	Simulated Public Speech	IntDiff	11.800	5.437	3	11.675	3
d	Academic Presentation	IntDiff	12.079	4.890	4	12.033	4
d	Read Speech	IntDiff	10.791	4.974	2	10.843	2
g	Dialogue	IntDiff	8.666	6.021	2	7.960	1
g	Simulated Public Speech	IntDiff	8.158	5.311	1	8.472	2
g	Academic Presentation	IntDiff	9.146	5.232	3	10.352	4
g	Read Speech	IntDiff	9.804	5.516	4	8.756	3

Since our analysis indicated that the extent to which reduction occurs across speech styles varied depending on a parameter of reduction, we employed another measurement to investigate it further. The intensity difference measure has been used for studies investigating the reduction of word-medial stops previously. Warner & Tucker (2011) found a smaller intensity difference for stronger reduction (approximant-like production), and the intensity difference becomes smaller for spontaneous speech than for read speech. Intensity difference was defined as the

difference between the minimum intensity of the target segment to the averaged maximum intensity of surrounding segments (Warner & Tucker, 2011). We predicted that a smaller intensity difference should be observed as speech becomes more spontaneous following this previous work on word-medial stops.

2.3.3 Intensity difference of voiced stops

Visual inspection of the distribution of intensity differences allowed us to find a few outliers skewing the distribution of data. As a result, we removed these outliers by excluding the stops with a intensity difference that was more than 2.5 standard deviations from the mean intensity difference for each phoneme. We then used the same modelling procedures as for the analysis of duration. The final models contained the same random intercepts and control variables as the models of duration analysis, indicating smaller intensity difference for shorter word duration and faster speech rate.

The mean intensity difference was smallest in /g/, second smallest in /b/, and largest in /d/. Pairwise comparisons between the phonemes revealed that the intensity difference of /d/ was larger than that of /b/ (t=3.129, $p<.005$) or that of /g/ (t=9.067, $p<.0001$), and the intensity difference of /d/ was larger than that of /g/ (t=5.425, $p<.0001$). For speech styles, the mean intensity difference was smallest in Dialogue, second smallest in Read Speech, third smallest in Simulated Public Speech, and largest in Academic Presentation. Pairwise comparisons among the speech styles revealed that the comparisons between Dialogue and Simulated Public Speech, between Dialogue and Academic Presentation, between Dialogue and Read Speech, and between Academic Presentation and Read Speech reached significance. In other words, the intensity difference was smaller in Dialogue than in Simulated Public Speech (t=3.056, $p<.005$), in Academic Presentation (t=5.231, $p<.0001$), or in Read Speech (t=3.652, $p<.005$), and the intensity difference was smaller in Read Speech than in Academic Presentation (t=3.050, $p<.01$).

As in the analysis of segment duration, we also investigated the interaction effect between the phonemes and speech styles, revealing that including the interaction improved the model (χ^2=17.6, $p<.01$). In order to further examine the

interaction effect, we divided the data into different phonemes and modeled them separately. Our final model for each phoneme included the same random intercepts and control variables as the models of duration analysis. Figure 2.3 shows the distribution of intensity differences of voiced stops for each speech style and phoneme. The figure reveals that distributional patterns of intensity difference differ from that of duration, and the patterns appear to be more consistent for the intensity difference than for the duration across phonemes. As a results, while duration covaries with the percentage of absent of a complete closure, the intensity difference appears to be independent of the percentage of absence of a complete closure.

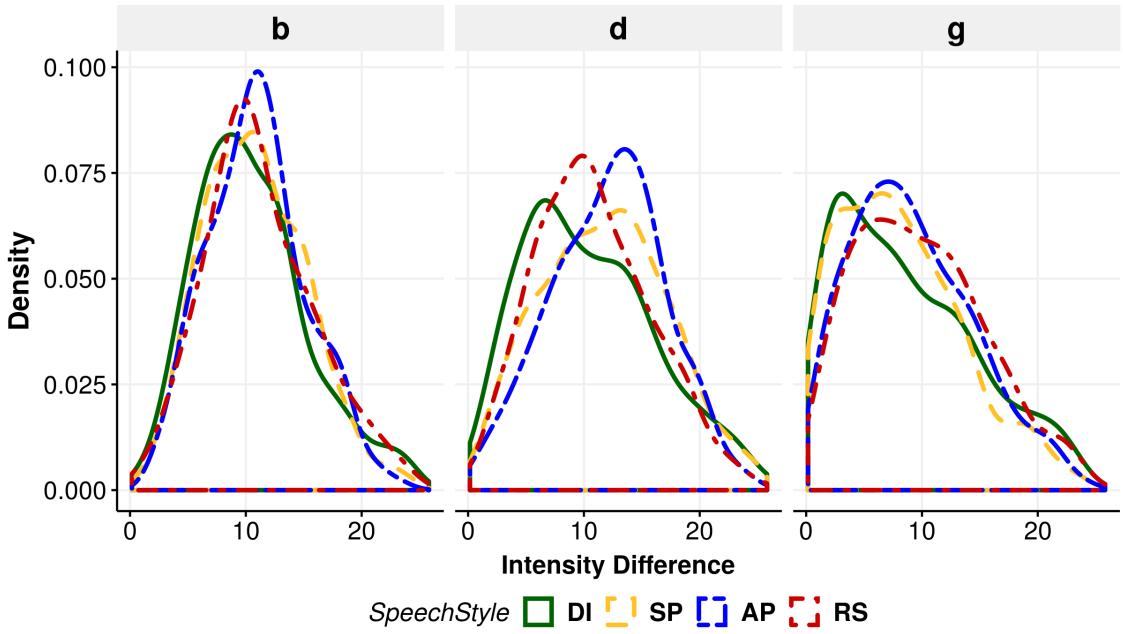

Figure 2.3: The distribution of intensity difference of voiced stops for each phoneme and speech style (DI: Dialogue; SP: Simulated Public Speech; AP: Academic Presentation; RS: Read Speech

Table 2.3 also illustrates an overview of intensity differences for each phoneme and speech style. For /b/, the mean intensity difference was smallest in Dialogue, second smallest in Simulated Public Speech, third smallest in Academic Presentation, and largest in Read Speech. The pairwise comparisons between the speech styles indicated that the intensity difference of /b/ was smaller in Dia-

logue than in Read Speech (t=2.911, $p<.01$). The mean intensity difference of /d/ was smallest in Dialogue, second smallest in Read Speech, third smallest in Simulated Public Speech, and largest in Academic Presentation (Table 2.3). Pairwise comparisons among the speech styles indicated that the comparisons between Dialogue and Simulated Public Speech, and between Dialogue and Academic Presentation reached significance. That is, the intensity difference of /d/ was smaller in Dialogue than in Simulated Public Speech (t=2.832, $p<.05$), or in Academic Presentation (t=3.394, $p<.005$). The mean intensity difference of /g/ was smallest in Simulated Public Speech, second smallest in Dialogue, third smallest in Academic Presentation, and largest in Read Speech. Pairwise comparisons across the speech styles demonstrated that the comparisons between Dialogue and Academic Presentation, between Simulated Public Speech and Academic Presentation, and between Academic Presentation and Read Speech reached significance. That is, the intensity difference of /g/ was smaller in Dialogue than in Academic Presentation (t=3.823, $p<.001$), and it was also smaller in Simulated Public Speech than in Academic Presentation (t=3.733, $p<.001$) or in Read Speech (t=2.603, $p<.05$). The estimated intensity difference was smallest in Dialogue, second smallest in Simulated Public Speech, third smallest in Read Speech and largest in Academic Presentation, and the comparisons between the smallest and largest intensity difference, between the second smallest and largest intensity difference, and between the third smallest and largest intensity difference reached significance. In other words, the intensity difference in Academic Presentation was larger than in all other speech styles.

Overall, the intensity difference measure reflected the speech style effect well. The intensity difference in Dialogue is smallest for all phonemes (although the mean difference in Dialogue is not smallest in /g/, the estimated values of intensity differences by the model revealed that the differences in Dialogue were smallest for all phonemes (Table 2.3)), and the majority of the pairwise comparisons between Dialogue and the other styles reached significance.

2.3.4 Nasals

Research has shown that nasals in Japanese are also realized in various forms in spontaneous speech (Arai, 1999; Vance, 2008). For example, in the case of the word-medial nasal in /tenisɯ/ 'tennis', the nasal and following /i/ are deleted, and the /e/ is both nasalized and lengthened. Therefore, it was realized as [tẽːsɯ] (Arai, 1999). Moreover, a syllable-final nasal /N/ will be homoroganic of the following segment, where /N/ becomes [m] before /p/ or /b/, or it becomes [n] before /t/, /d/, or /ɾ/, leading to a wider range of allophonic variations (Vance, 1987, 2008). While such assimilation has been studied in the context of phonology extensively, more research is necessary from the acoustic-phonetic point of view (Cutler & Otake, 1998; Mizoguchi, 2019). We expect a high phonetic variability for /N/ (because we used phonological transcriptions), but its duration should be independent of the allophonic variations because /N/ bears a mora by itself and duration of all morae are similar (Kawahara, 2017; Vance, 1987, 2008). Furthermore, while we have observed instances of phonetic reduction of nasals, there are fewer studies that discuss the distribution and degree of such reduction across various styles of speech (Arai et al., 2007; Nakamura et al., 2008). We therefore employ similar measures as we used to measure reduction in word-medial voiced stops and investigate the interplay between reduction and style of speech for word-medial nasals in Japanese.

2.3.5 Duration of nasals

As in the case of the analysis of voiced stops, we used linear mixed-effects modelling to compare the duration of nasals across phonemes and speech styles. We used the same modelling procedures as the analysis of voiced stops, including the same random intercepts and control variables. Our final models comprised the two random intercepts, with the duration of words, speech rate, and type of mora as control variables, indicating shorter word duration, faster speech rate, and unaccented mora for shorter segment duration.

The mean duration was longest in /N/, second longest in /m/, and shortest

in /n/. Pairwise comparisons between the phonemes revealed that all of them were different from each other; the duration of /N/ was longer than that of /m/ (t=13.084, $p<.0001$) or /n/ (t=25.263, $p<.0001$), and the duration of /m/ was longer than that of /n/ (t=12.391, $p<.0001$). For speech styles, the mean duration was longest in Read Speech, second longest in Dialogue, third longest in Simulated Public Speech, and shortest in Academic Presentation. Pairwise comparisons between the speech styles indicated that the comparisons between Dialogue and Read Speech, between Simulated Public Speech and Read Speech, and between Academic Presentation and Read Speech reached significance. That is, the duration was longer in Read Speech than in Dialogue (t=5.382, $p<.0001$), in Simulated Public Speech (t=3.287, $p<.01$), or in Academic Presentation (t=4.491, $p<.0001$).

Similar to the analysis of voiced stops, we also investigated the interaction effect between the phonemes and speech styles, revealing that including the interaction improved the model (χ^2=47.714, $p<.0001$). For further examination, we divided the data into the different phonemes and modeled them separately. Our final model for each phoneme included the same random intercepts and control variables except for /m/ and /N/ excluding the type of mora. Figure 2.4 shows the distribution of duration of nasals for each speech style and phoneme. The figure reveals that the distributional pattern of duration among speech styles appears to be somewhat consistent for the /m/ and /n/ phonemes except for /N/ showing a longer duration and wider range of durational variation.

Table 2.4 illustrates an overview of duration for each phoneme and speech style. The mean duration of /m/ was longest in Read Speech, second longest in Simulated Public Speech, third longest in Dialogue, and shortest in Academic Presentation. Pairwise comparisons across the speech styles indicated that none of the comparisons reached significance. The mean duration of /n/ was longest in Read Speech, second longest in Dialogue, third longest in Simulated Public Speech, and shortest in Academic Presentation. Pairwise comparisons across the speech styles indicated that none of the comparisons reached significance. The mean duration of /N/ was longest in Read Speech, second longest in Simulated Public Speech, third longest in Dialogue, and shortest in Academic Presentation. In contrast to

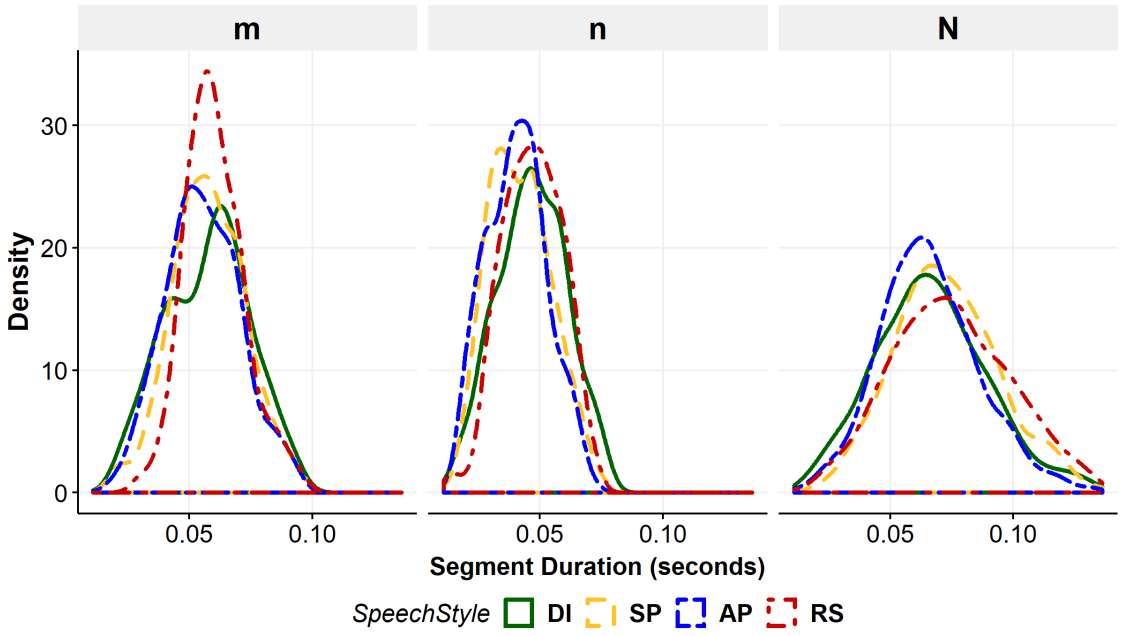

Figure 2.4: The distribution of duration of nasals for each phoneme and speech style (DI: Dialogue; SP: Simulated Public Speech; AP: Academic Presentation; RS: Read Speech)

the other two phonemes, pairwise comparisons between the speech styles indicated that the comparisons between Dialogue and Read Speech, and between Academic Presentation and Read Speech reached significance; that is, the duration was longer in Read Speech than in Dialogue (t=4.804, $p<.0001$), or in Academic Presentation (t=4.045, $p<.0001$).

In the present analysis, we found that the duration of nasal is longest in Read Speech for all phonemes. However, none of the pairwise comparisons reached significance for /m/ and /n/, suggesting that the overall effect of speech style and interaction between speech style and phoneme resulted solely from /N/.

2.3.6 Intensity difference of nasals

We also employed intensity difference measure as we used for the analysis of voiced stops. We employed the same modelling procedures as the analysis of duration, including the same random intercepts and control variables. The final model for the analysis of phonemes included the two random intercepts, with the

Table 2.4: Means and standard deviations (SD) of duration (in seconds) and intensity difference (IntDiff in dB), as well as estimated values of duration and intensity difference by the models, for phonemes and speech styles. For Style, Dialogue; Simulated Public Speech; Academic Presentation; Read Speech. In Rank and Est.Rank columns, 1 indicates the shortest duration and smallest intensity difference and 4 indicates the longest duration and largest intensity difference.

Segment	Style	Parameter	Mean	SD	Rank	Est.Value	Est.Rank
m	Dialogue	Duration	0.0574	0.0170	2	0.0553	1
m	Simulated Public Speech	Duration	0.0577	0.0148	3	0.0575	3
m	Academic Presentation	Duration	0.0555	0.0152	1	0.0561	2
m	Read Speech	Duration	0.0602	0.0122	4	0.0582	4
n	Dialogue	Duration	0.0468	0.0140	3	0.0458	4
n	Simulated Public Speech	Duration	0.0420	0.0127	2	0.0435	1
n	Academic Presentation	Duration	0.0406	0.0122	1	0.0423	2
n	Read Speech	Duration	0.0469	0.0120	4	0.0449	3
N	Dialogue	Duration	0.0667	0.0229	2	0.0695	2
N	Simulated Public Speech	Duration	0.0729	0.0213	3	0.0719	3
N	Academic Presentation	Duration	0.0659	0.0203	1	0.0686	1
N	Read Speech	Duration	0.0745	0.0239	4	0.0752	4
m	Dialogue	IntDiff	2.677	1.960	1	3.134	1
m	Simulated Public Speech	IntDiff	3.452	2.146	3	3.671	3
m	Academic Presentation	IntDiff	3.733	2.079	4	4.083	4
m	Read Speech	IntDiff	3.184	1.938	2	3.582	2
n	Dialogue	IntDiff	2.220	1.608	1	2.516	1
n	Simulated Public Speech	IntDiff	2.706	1.686	3	2.932	3
n	Academic Presentation	IntDiff	3.243	1.942	4	3.400	4
n	Read Speech	IntDiff	2.594	1.634	2	2.764	2
N	Dialogue	IntDiff	9.120	4.849	3	9.120	3
N	Simulated Public Speech	IntDiff	9.002	5.373	2	8.948	2
N	Academic Presentation	IntDiff	8.507	4.387	1	8.895	1
N	Read Speech	IntDiff	10.173	4.967	4	10.603	4

duration of words and speech rate as control variables, and the final model for the analysis of speech styles remained all the control variables. These variables indicated that smaller intensity difference was associated with higher frequency words, shorter word duration, faster speech rate, and unaccented mora.

The mean intensity difference was smallest in /n/, second smallest in /m/, and largest in /N/. Pairwise comparisons between the phonemes revealed that the comparisons between /m/ and /N/, and /n/ and /N/ reached significance. That is, the intensity difference of /N/ was larger than that of /m/ (t=29.217 $p<.0001$) or /n/ (t=28.720, $p<.0001$). As for speech styles, the mean intensity difference was smallest in Simulated Public Speech, second smallest in Dialogue, third smallest in Academic Presentation, and largest in Read Speech. Pairwise comparisons among the speech styles indicated that the comparisons between Dialogue and Read Speech, and between Simulated Public Speech and Read Speech reached significance. In other words, the intensity difference was larger in Read Speech than in Dialogue (t=7.008, $p<.0001$) or in Simulated Public Speech (t=3.798, $p<.0005$).

Similarly, we also investigated the interaction between the phonemes and speech styles, revealing that including the interaction improved the model (χ^2=107.33, $p<.0001$). We divided the data into different phonemes and modeled them separately. Our final model for each phoneme included the two random intercepts and following control variables: the type of mora for /m/, no control variables for /n/, and word duration and speech rate for /N/. Figure 4 shows the distribution of intensity difference of nasals for each speech style and phoneme. The figure reveals that the distribution of intensity difference across speech styles appears to be similar between /m/ and /n/, but /N/ depicts a wider range of intensity difference.

Table 2.4 illustrates the overview of intensity difference for each phoneme and speech style. For /m/, the mean intensity difference was smallest in Dialogue, second smallest in Read Speech, third smallest in Simulated Public Speech, and largest in Academic Presentation. The pairwise comparisons between the speech styles indicated that the comparison between Dialogue and Academic Presentation reached significance. That is, the mean intensity difference of /m/ was smaller in

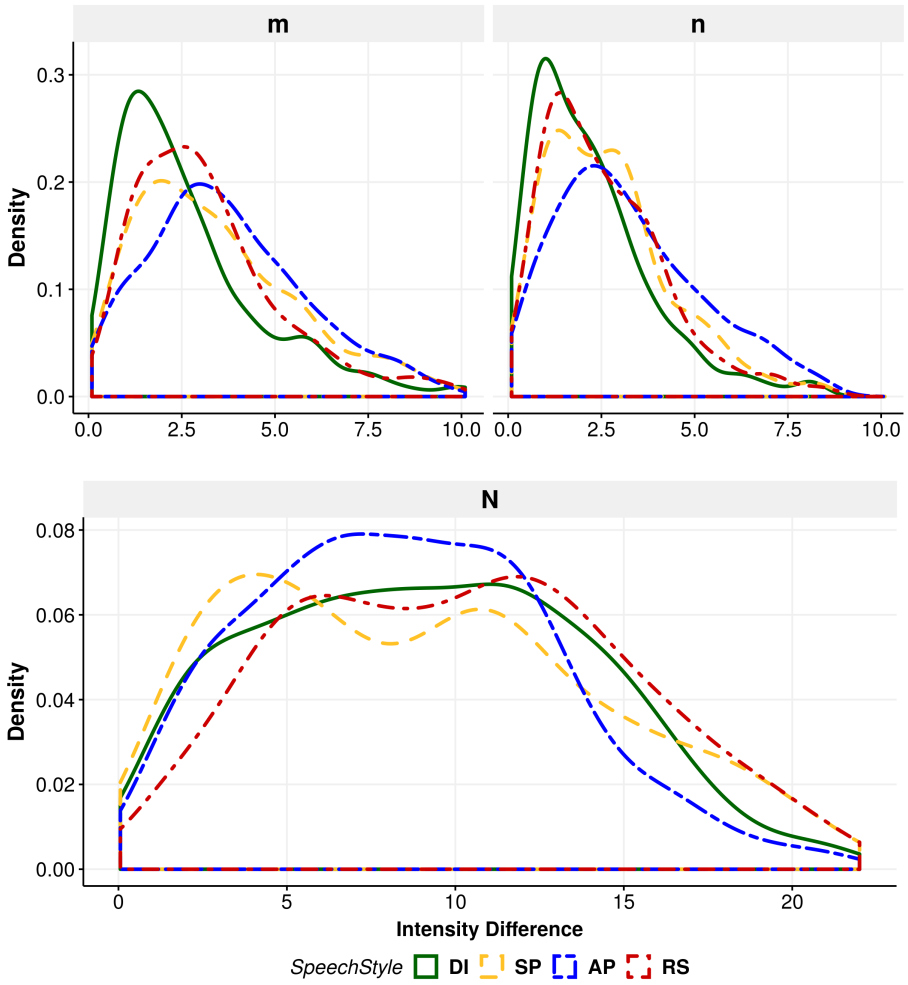

Figure 2.5: The distribution of intensity difference of nasals for each phoneme and speech style (DI: Dialogue; SP: Simulated Public Speech; AP: Academic Presentation; RS: Read Speech)

Dialogue than in Academic Presentation (t=3.682, $p<.001$). The mean intensity difference of /n/ was smallest in Dialogue, second smallest in Read Speech, third smallest in Simulated Public Speech, and largest in Academic Presentation (Table 2.4). The pairwise comparisons between the speech styles revealed that the comparisons between Dialogue and Academic Presentation, and between Read Speech and Academic Presentation reached significance. In other words, the mean intensity difference of /n/ was larger in Academic Presentation than in Dialogue (t=3.964, $p<.0005$) or in Read Speech (t=2.863, $p<.05$). The mean intensity difference of /N/ was smallest in Academic Presentation, second smallest in Simulated Public Speech, third smallest in Dialogue, and largest in Read Speech. Pairwise comparisons among the speech styles revealed that the comparisons between Dialogue and Read Speech, between Simulated Public Speech and Read Speech, and between Academic Presentation and Read Speech reached significance. In other words, the mean intensity difference of /N/ was larger in Read Speech than in Dialogue (t=5.586, $p<.0001$), in Simulated Public Speech (t=3.884, $p<.0005$), or in Academic Presentation (t=4.655, $p<.0001$).

In summary, the effect of speech style was well reflected in the intensity difference measure. The intensity difference in Dialogue was smallest for /m/ and /n/ (although it was second largest for /N/), and the pairwise comparisons between the smallest and largest intensity difference reached significance for all phonemes. The different pattern of speech style effect for /N/ could be due to the assimilation process that /N/ undergoes and the longer temporal space that /N/ holds. Future research is needed for a separate analysis for /N/ taking into account the different assimilation environments for each style of speech.

2.4 Discussion

The aim of this study was to examine phonetic reduction in word-medial voiced stops and word-medial nasals in Japanese and to describe how that reduction occurs across speech styles in Japanese. Specifically, we compared the distribution and degree of phonetic reduction across various styles of speech to investigate the

effect of speech styles on the realization of word-medial stops and word-medial nasals. In the section follows, we will illustrate how comparable the acoustic results of reduction are between Japanese and English across speech styles. First, to summarize our results, we found that the interaction between the effect of reduction and speech style is dynamic.

For the voiced stops, when using the presence or absence of a complete closure as a parameter of reduction, overall we found an effect of speech style and phoneme, as well as their interaction, indicating less reduction for the least spontaneous speech, Read Speech, than for the most spontaneous speech, Dialogue. This result is in line with what has been found in Warner & Tucker (2011) demonstrating greater reduction for more a spontaneous speech style in English (their analysis included both voiced and voiceless stops). They used two types of read speech: word-list reading and story reading, and one type of spontaneous speech: a telephone conversation with a friend or family member. The difference in the degree of spontaneity is likely greater among their speech styles, particularly between the telephone conversation and word list reading or story reading, than among our speech styles. Moreover, while it has been claimed that alveolar consonants are prone to reduction more than other consonants (Barry & Andreeva, 2001), Warner & Tucker (2011) and our study indicates that /g/ was reduced the most among voiced stops.

When we used duration as a parameter of reduction, an effect of speech style and phoneme, as well as their interaction, also appeared, which is in line with the findings of Warner & Tucker (2011). However, the direction of the effect differs between our findings and the Warner & Tucker (2011) findings. Whereas Warner & Tucker (2011) found greater reduction for more spontaneous speech (shorter duration for more spontaneous speech), we found the least reduction in the most spontaneous speech (Dialogue) for all phonemes. Warner & Tucker (2011), as well as Tucker (2007), argue that duration may not be an adequate parameter of reduction by itself due to other factors influencing segment duration, such as the lengthening of stressed and shortening of unstressed syllables. Furthermore, they also indicate that the effect of speech style was consistently smaller than that of

phoneme and that the difference between a slow and fast talker can be as large as the difference between the different styles of speech. The speech rate difference could be a reason that our result differs from Warner & Tucker (2011). Our Japanese corpus comprises speech styles that lean towards fast speech, possibly leading to shorter segment duration. While the overall segment duration is similar between Dialogue in our study and the spontaneous conversation in Warner & Tucker (2011), the overall segment duration of Academic Presentation, Simulated Public Speech, and Read Speech are relatively shorter than that of the story reading and word list reading in Warner & Tucker (2011). We calculated speech rate of our data by the number of vowels per second (Dilts, 2013). The rate was fastest in Academic Presentation, second fastest in Simulated Public Speech, third fastest in Dialogue, and slowest in Read Speech. The possible reasons that our styles show faster rate are as follows. First, Academic Presentation is prepared speech. Being familiar with what to say and practicing how to deliver the talk could lead speakers to speak quickly. Second, in the case of Read Speech, readers were familiar with the text because they read the transcriptions of their academic presentation. Another possible reason that our result differs from Warner & Tucker (2011) could be that our study kept potential confounding factors for segment duration constant (e.g., word duration, accented or unaccented mora, and speech rate). It may be the case that the speech style effect in Warner & Tucker (2011) may change when these factors are held constant, as approximated articulation does not necessarily have to lead to shortening of segments (Turnbull, 2018). Finally, both studies find interaction effects between speech style and phoneme. In our analysis, one pairwise comparison that reached significance was between Dialogue (longest) and Read Speech (shortest) for /b/. In Warner & Tucker (2011) for voiced stops, /b/ shows the effect of speech style but /d/ did not, and /g/ was excluded in their analysis (due to lack of data in one of the conditions). That is, /b/ is the only segment that manifests the effect of speech style for both studies.

Using intensity difference as a parameter of reduction, we found the most consistent effects of speech style. The intensity difference results followed the predicted reduction pattern, where the most spontaneous speech demonstrates the

greatest reduction. However, the pairwise comparisons between speech styles for each phoneme revealed differences only between the smallest (Dialogue) and largest intensity difference (Academic Presentation or Read Speech). If speech styles are represented on a continuum, then the distance between the four speech styles is variable where some are closer to others. While there may be an ordered difference between these speech styles, not all measures may find differences. This could be because, as stated in the descriptions of CSJ, the stylistic difference between Academic Presentation and Simulated Public Speech might be smaller than between Dialogue and Read Speech. The descriptions of CSJ can be found at https://pj.ninjal.ac.jp/corpus_center/csj/manu-f/overview.pdf (This document is written in Japanese). The result of our study is in agreement with Warner & Tucker (2011) and Maekawa (2005), in that we all demonstrate that more spontaneous speech contains greater reduction. While the values of intensity difference for Dialogue in our study and for spontaneous speech type in the Warner & Tucker (2011)'s study were comparable, the values of intensity difference in read speech in their study appeared to be relatively higher than that of intensity difference in Read Speech from our study. This could be because of the tasks given to speakers. While speakers in their study only read their story once, speakers in our study read the transcriptions of their talk (Academic Presentation), meaning that they were familiar with the text and might have practiced producing it many times. This could lead speakers to speak quickly and/or approximate their speech articulation.

For nasals, although the overall effect of speech style and the interaction effect between speech style and phoneme resulted solely from /N/ when we used duration as a parameter of reduction, all phonemes demonstrated a tendency of the predicted reduction pattern, the least spontaneous speech indicating the smallest reduction. The duration of /N/ was longer in Read Speech than in Academic Presentation or in Dialogue. The reason that the pairwise comparisons only for /N/ reached significance is possibly due to the extra temporal space that /N/ bears. Similar to the results of voiced stops, the effect of speech style was better reflected in the intensity difference. The intensity difference of /n/ and /m/ in Dialogue

was smaller than in the other styles, showing the predicted reduction patterns, the most spontaneous speech demonstrates the greatest reduction. For /N/, the intensity difference was larger in Read Speech than in Dialogue, in Simulated Public Speech, or in Academic Presentation, also indicating the predicted pattern, the least spontaneous speech demonstrates the smallest reduction. Similar to the results in the analysis of duration, the pattern of intensity difference between speech styles for /N/ differs from /m/ and /n/ probably due to the assimilation process that /N/ undergoes (Vance, 1987, 2008). Further analysis of /N/ considering the different assimilation environments for each style of speech is necessary to better understand reduction for this segment. In short, this result extends the findings of Warner & Tucker (2011) in a way that the role of speech style on the realization of nasals is reflected in a similar way to that of stops and flaps, particularly when intensity differences are utilized as a measure of reduction.

In short, there seems to be a gradual effect of reduction across speech styles specifically when using intensity difference as a measure of reduction. The effect illustrates a tendency of predicted reduction patterns, the most spontaneous speech demonstrates the greatest reduction, and the least spontaneous speech exhibits the smallest reduction. However, we did not find a clear gradual effect as Maekawa (2005) suggested. That is, although we found the tendency of the predicted reduction patterns across the speech styles, the overall effect was more categorical; Dialogue (most spontaneous) shows the greater reduction than Read Speech (the least spontaneous). This could be because the effect of speech style is not as strong as expected especially when other confounding factors are held constant. Overall, when we employ an intensity difference as a measure of reduction, the acoustic results of reduction and their distributional patterns across speech styles for voiced stops are most comparable between Japanese in the present study and English in Warner & Tucker (2011), and nasals also indicate comparable reduction patterns. These results extend the findings of Barry & Andreeva (2001) that there are language-independent patterns of reduction regardless of language rhythm types (syllable-time in English vs. mora-time language in Japanese), That is, although there is a limited structural basis of consonant reduction in Japanese stops (i.e.,

consonant clusters), voiced stops (and possibly nasals) in Japanese and English appear to display a similar distribution and degree of reduction across different speech styles.

Our results might have been different if we had used different speech styles because speech style is likely multi-dimensional (Tucker, 2007). While our focus was on a spontaneous vs. read speech type, the situation where the speech is produced (e.g., comfortable vs. uncomfortable) could also be relevant dimension. Whereas spontaneous speech in Warner & Tucker (2011) is considered comfortable casual speech (a telephone conversation with a friend or family member), Dialogue in our study is considered uncomfortable casual speech to some extent because although it was carried out in a relatively casual manner, the two speakers did not know each other (participant and experimenter), which might have made their conversation somewhat formal and slower. The formality could also be another dimension to consider (despite the fact that it could be intertwined with the careful/casual distinction). Academic Presentation in our study is considered formal, and Read Speech in our study is also considered formal because the texts they read was their academic presentation. Story reading in Warner & Tucker (2011) could be either formal or informal depending on the content of the reading. For example, how readers express their readings could differ between reading a story about political debates (formal) or their hobbies (informal).

Illustrating the dynamics of speech style is necessary to account for processes that occur in speech production because speakers vary their production depending on a given context (e.g., Lindblom, 1990). Importantly, a casual, spontaneous speech style is what we encounter the most and it substantially differs from a careful, read speech style that is often used in speech production studies. We believe that it is important to investigate how speech style differences influence the production of speech because it could lead us to new insights and questions on the underlying mechanisms of speech production that cannot be revealed by research only using a careful, laboratory speech style (Ernestus & Warner, 2011; Tucker & Ernestus, 2016).

2.5 Conclusion

In this study, we investigated the effect of speech style on the realization of word-medial voiced stops and word-medial nasals and compared our acoustic results to the results in Warner & Tucker (2011). The effect of speech style appeared to be gradual but it did not indicate a clear gradual effect as Maekawa (2005) suggested; the most spontaneous speech showed greater reduction than the least spontaneous speech. The overall acoustic results of reduction and their distribution patterns across speech styles are similar between Japanese and English. Nasals also produce comparable reduction patterns. This result supports the hypothesis in Barry & Andreeva (2001) that there are cross linguistic patterns of reduction across languages regardless of rhythmic types.

Chapter 3

The effect of phonological-orthographic consistency in the recognition of reduced speech for L1 speakers: Evidence from pupillometry

3.1 Introduction

Inconsistencies between the way in which words are pronounced and spelled have been shown to affect the recognition of spoken words (Pattamadilok, Kolinsky, Ventura, Radeau & Morais, 2007; Veivo & Järvikivi, 2013; Ziegler & Ferrand, 1998; Ziegler et al., 2004). Research has shown that auditory lexical decisions are slower and less accurate for inconsistent words whose rhyme can be spelled in multiple ways, such as /-ip/ in "leap" or "keep" (Ziegler & Ferrand, 1998). Subsequent research has replicated this phenomenon, referred to as the phonological-orthographic (P-O) consistency effect, in rhyme detection tasks, picture naming, and spoken word recognition with visual world eye-tracking (Salverda & Tanenhaus, 2010; Veivo et al., 2016), but has failed to replicate it in an auditory naming task for both existing and novel words (Rastle, McCormick, Bayliss & Davis, 2011; Ziegler et al., 2004).[1]

Previous research has investigated the P-O consistency effect using carefully

[1]This chapter was originally written as a brief research report. A comprehensive literature review and further discussion of results are provided in Chapter 4 and 5.

pronounced words. Spoken language is, however, highly variable, particularly in casual every-day speech. This variability often results from phonetic reduction as in the realization of words with approximated articulation, resulting in deletion and/or incomplete articulation of segments (Ernestus & Warner, 2011; Warner & Tucker, 2011). For example, *yesterday* /jɛstəˇdeɪ/ may be pronounced as [jɛʃeɪ] (Tucker, 2007). On the one hand, this implies that the inconsistency between the way *yesterday* is pronounced and spelled should be greater for the casual pronunciation ([jɛʃeɪ] and *yesterday*) than for the careful pronunciation (/jɛstəˇdeɪ/ and *yesterday*). On the other hand, Mitterer & Reinisch (2015) have argued that the effect of orthography, including the P-O consistency effect, does not play an important role in the recognition of casual speech, suggesting that there is no effect of P-O consistency for casually pronounced words. Importantly, these reduced forms are recognized less efficiently than canonical/unreduced counterparts despite the fact that reduced forms occur more frequently than canonical/unreduced ones (e.g., Arai et al., 2007; Ernestus et al., 2002; Tucker, 2007, 2011; van de Ven et al., 2011). Some researchers have argued that the reason that unreduced forms are easier to process than reduced counterparts could be because of the orthography, particularly the consistent relationship between the unreduced pronunciation and its orthographic form (Racine et al., 2014; Ranbom & Connine, 2007, 2011; Viebahn et al., 2018), suggesting that phonological and orthographic consistency should play an important role in the processing of reduced forms. In the present study, we investigate how P-O consistency interacts with phonetic reduction using Go-NoGo and delayed naming tasks. Specifically, we compare the time course of the P-O consistency effect between reduced and unreduced pronunciations (casual .vs careful) of Japanese words.

While several languages have been tested on the P-O consistency effect (Chen et al., 2016; Hino et al., 2017; Ventura et al., 2004; Ziegler & Ferrand, 1998), research on this effect in Japanese is very limited. Japanese P-O consistency is measured using the P-O consistency index, which is calculated based on the number and frequency of phonological and orthographic neighbours of a target word (Hino et al., 2017). After identifying all phonological neighbours of a target word, the phono-

logical neighbours are classified into two types: orthographic friends and orthographic enemies (Jared, McRae & Seidenberg, 1990; Ziegler, Muneaux & Grainger, 2003). If the phonological neighbour is also an orthographic neighbour of the target word, it is categorized as an orthographic friend; if not, it is categorized as an orthographic enemy. Following this classification, the frequencies of the target word and orthographic friends are summed and divided by the sum of the frequencies of the target word, orthographic friends, and orthographic enemies. The P-O consistency index ranges from 0 to 1, with 0 indicating low consistency and 1 indicating high consistency. Hino et al. (2017) found that auditory lexical decisions are slower for low consistency Japanese words as compared to high consistency words.

As in other studies, Hino et al. (2017) also used unreduced Japanese words. In the current study, we focus on the reduction of word-medial nasals and voiced stops, given that Arai (1999) and Mukai & Tucker (2017) have demonstrated that both consonant types show various forms of reduction. For example, in the case of the word-medial nasal in /tenisɯ/ 'tennis', the nasal and following /i/ are deleted, and the /e/ is both nasalized and lengthened; the word is then realized as [tẽːsɯ] (Arai, 1999). For word-medial voiced stops, articulation is often approximated due to the lack of full oral closure, and in the extreme case, the consonant is deleted: /dɑigɑkɯ/ → [dɑiɣɑkɯ] → [dɑiɑkɯ] (Arai, 1999; Mukai & Tucker, 2017).

In the present study, we compare the time course of the P-O consistency effect between reduced and unreduced forms of Japanese words as indicated by pupil dilation. The pupil has been shown to respond to physiological arousal during cognitive tasks (Beatty, 1982). Pupil dilation correlates with the amount of cognitive effort induced by tasks (Zekveld, Kramer & Festen, 2010), and it has been utilized as an index of cognitive load (Kahneman & Beatty, 1966) and applied to a variety of psycholinguistic studies (Geller, Still & Morris, 2016; Kuchinke, Vo, Hofmann & Jacobs, 2007; Papesh & Goldinger, 2012; Porretta & Tucker, 2019; Zekveld & Kramer, 2014). Pupillometry offers a reliable method to examine allocations of cognitive resources imposed by different variables in speech comprehension (Laeng et al.,

2012). For our experiments, 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).

Using pupillometry, we examine the time course of the effect of P-O consistency with reduced and unreduced Japanese words. If P-O consistency affects reduced and unreduced forms differently, we should observe an interaction between the effect of reduction and P-O consistency, suggesting that either reduction creates an additional sound-to-orthography mismatch or that the P-O consistency effect does not affect reduced forms as much as unreduced forms as Mitterer & Reinisch (2015) have suggested.

3.2 Method

We conducted a Go-NoGo and delayed naming task. As in the study by Perre et al. (2011), participants respond to particular stimuli (Go) but they do not respond to a different set of stimuli (NoGo) in the Go-NoGo task. In our task, both reduced and unreduced forms of Japanese words served as non response stimuli (NoGo) while pure tones are response stimuli (Go). In the delayed naming task, participants hear a target word and wait for a response signal. After the signal, they repeat what they have heard. In our task, participants hear a Japanese word and wait until they hear a pure tone. They then repeat what they have heard.

3.2.1 Participants

Thirty-eight native speakers of Japanese (female, n = 16) ranging in age from 18 to 25 years old (M = 19.7, SD = 1.69) were recruited at Nagoya University in Japan. All participants reported normal or corrected-to-normal vision and hearing. All participants performed both tasks and half of them performed the Go-NoGo task first and the other half did the delayed naming task first.

SegmentType	Reduction	WordDuration	SegmentDuration	SpeechRate	MeanWordPitch	IntDifference
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	$p < 0.001$	$p < 0.001$	$p < 0.001$	$p < 0.001$	$p = 0.178$

Table 3.1: Mean acoustic values of stimuli in reduced and unreduced forms. Both word and segment durations are represented in seconds. Speech rate was measured as the number of vowels per second. p *values indicate the probability that the difference between the two forms were significantly different from 0.*

3.2.2 Materials

We selected 226 four-mora and two-logograph words (lists of the words are available on Education and Research Archive: https://doi.org/10.7939/r3-60xn-qd28). We used the Balanced Corpus of Contemporary Written Japanese (Maekawa et al., 2014) to obtain word frequency information and calculate the P-O consistency index for each target word. For the calculation of P-O consistency, phonological neighbours were defined as words that differ by a single mora from the target word, and orthographic neighbours were defined as words that differ by a single character from the target word (Fushimi et al., 1999). All words contain a word-medial nasal and/or voiced stop and were recorded in both reduced and unreduced forms by a female native Japanese speaker (452 item in total). We instructed the speaker to produce the words clearly for unreduced forms and casually (spontaneous speech like) for reduced forms. The speaker produced multiple tokens of both forms, and we selected the most natural sounding tokens as stimuli. For presentation purposes, we normalized the amplitude of the words. Table 1 indicates the acoustic properties of reduced and unreduced forms. For the voiced stops, we defined the intensity difference as the difference between the minimum intensity of the target segment to the averaged maximum intensity of surrounding segments (Warner & Tucker, 2011). Overall, reduced forms have shorter duration, faster speech rate, lower mean pitch, and smaller intensity difference. The differences between the two forms reached statistical significance for all properties except the intensity difference.

We created four lists for each task and each list contained 150 items (5 practice

words, 113 target words (reduced and 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). Similar to Perre et al. (2011), we employed a 500-ms-long pure tone as non-target items for the Go-NoGo task and manipulated the ratio between the target and non-target trials as 70% and 30%. The target words were counterbalanced across both reduction and task, so that none of the participants heard the same word twice.

3.2.3 Apparatus and procedure

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

In the Go-NoGo task, participants looked at a fixation cross presented at the centre of the screen on a gray background for 1500 ms and heard either a Japanese word or a pure tone as they continued looking at the fixation cross. They then responded to the pure tone by pressing a button on a Microsoft Side Winder gamepad or did not respond to the Japanese word. The fixation cross disappeared 2000 ms after the onset of Japanese words or after the button press trigged by the 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 (Papesh & Goldinger, 2012).

In the delayed naming task, participants looked at a fixation cross presented at the centre 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 500ms-long pure tone. They then repeated what they had heard. The fixation cross disappeared 2000 ms after the onset of Japanese words. 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 for both tasks, the practice items were provided at the beginning of sessions to familiarize the participants with the task. The target and non-target items were randomly assigned to each trial by the software. Participants took a brief break every 29 trials and each task lasted approximately 45 minutes.

3.2.4 Preprocessing pupil size data

We removed eye-blinks and their artifacts, 50 samples before and after the blinks and 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. Four participants were excluded from our data due to excessive blinks and their artifacts (more than 50% of trials contained more than 30% of eye-blinks and their artifacts). 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 location as a control variable. 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 subtract 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, Fabius, Heusden & der Stigchel, 2018). 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 for the Go-NoGo task and from the onset of stimuli to 2500 ms after the onset for the delayed naming task.

Data were cleaned and checked visually on a trial-by-trial basis to detect unexpected deviations (Winn, Wendt, Koelewijn & Kuchinsky, 2018). The trials that contained excessive blinks and their artifacts (more than 30% of the trial) were

excluded. Additional trials were excluded when the peak latency was shorter than 400 ms, the peak dilation was smaller than 0 or bigger than 400. In total, we excluded 19.7% of data in the Go-NoGo task and 20.7% of data in the delayed naming task. The data is available on Education and Research Archive (https://doi.org/10.7939/r3-60xn-qd28).

3.2.5 Analysis and results

For both tasks, we employed generalized additive mixed-modeling (GAMM) (Hastie & Tibshirani, 1990; Wood, 2006) because GAMM allows us to model non-linear relationships, as well as linear relationships, between a response variable and predictor variables (Porretta, Tremblay & Bolger, 2017; Sóskuthy, 2017; Wieling et al., 2016). This was important as we expected pupil size to fluctuate over time (van Rij et al., 2019). Additionally, GAMM allowed us to control serial dependency in time-series data, namely, autocorrelation (Baayen et al., 2017). All analyses were performed in R (R Development Core Team, 2018) using "mgcv" (Wood, 2017) and "itsadug" (van Rij, Wieling, Baayen & van Rijn, 2017) packages.

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 (0 - 1), Reduction (reduced or unreduced form) and Time (in milliseconds) as predictor variables. We also used, as control variables, Word Duration (in seconds), Target Segment (word-medial 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), Word Frequency (Log-transformed), Number of Homophones (Z-transformed), and Logged Number of Phonological Neighbours. The correlation between Logged Word Frequency and P-O Consistency was weak ($r = 0.27$), but Logged Number of Phonological Neighbours was negatively correlated with P-O Consistency ($r = -0.66$); high consistency words tended to have a small number of phonological neighbours. These two variables were, therefore, not included in the same model.

Go-NoGo task

We inspected the aggregated raw pupil dilation data prior to fitting models. Figure 3.1 illustrates the grand average of pupillary responses over time for reduced and unreduced forms from -1500 ms to 2000 ms in the Go-NoGo task. 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.

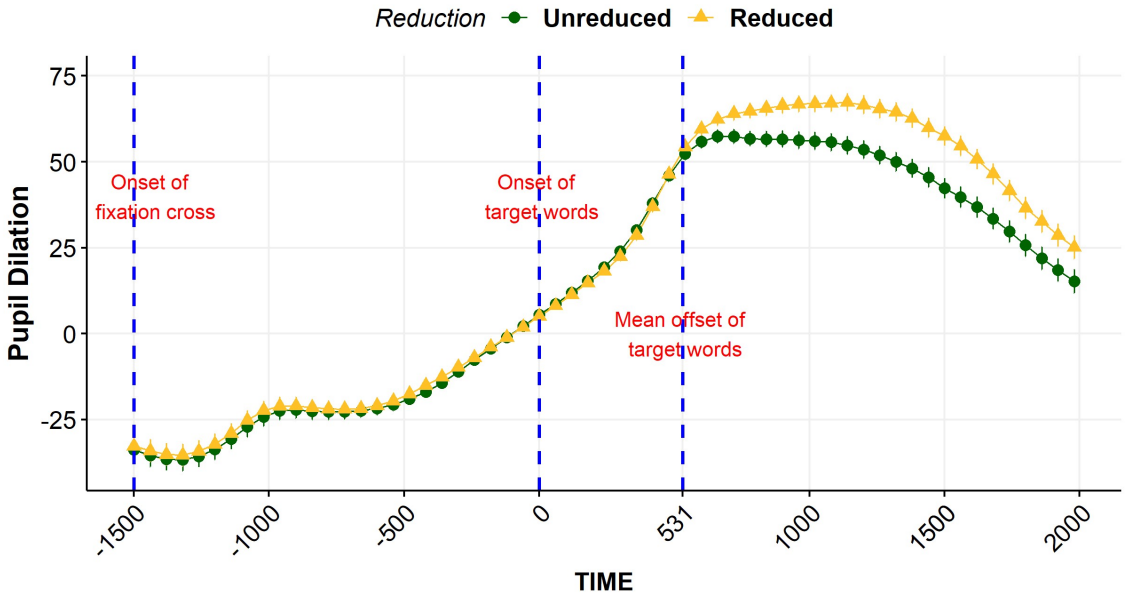

Figure 3.1: The grand average of pupillary responses over time for reduced and unreduced word forms in the Go-NoGo task. The vertical dotted line at -1500 ms indicates the onset of the fixation cross, the line at 0 ms indicates the onset of stimuli and the line at 531 ms indicates the mean offset of stimuli.

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). Using a smooth function, Pupil Dilation was fitted as a function of P-O Consistency, Reduction, and Time with Word Duration, Target Segment, Baseline Pupil Size, Pupil Gaze Coordinate X and Y, Trial Index, Logged Word Frequency, and Z-transformed Number of Homophones as control variables. Using a tensor product, we included the three-way interaction between Time, P-O consistency index, and Reduction. We employed a backwards stepwise elimination procedure for fixed effects and a forward fitting procedure

for random effects to fit the optimal model (Matuschek et al., 2017).

For fixed effects, we eliminated Target Segment, Logged Word Frequency, and Z-transformed Number of Homophones. Word Duration and Baseline Pupil Size were refitted without a smooth function because their effects were linear. For random effects, we included two factor smooths: ParIDConsis (unique combination of Participant ID and P-O Consistency) 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 to reflect participant-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 (Wieling, 2018). We also included an AR-1 correlation parameter at the value of 0.98 to address autocorrelation and fitted the model with the scaled-t family in order for residuals to be normally distributed (Meulman, Wieling, Sprenger, Stowe & Schmid, 2015; 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 parametric coefficients indicate that the overall pupil dilation is greater for reduced forms than for unreduced forms ($t = 5.453, p < 0.0001$), the overall pupil dilation is smaller for larger baseline pupil sizes ($t = -7.696, p < 0.0001$), and that the overall pupil dilation becomes greater as word duration increases ($t = 3.781, p < 0.0001$). The smooth terms reveal the significance of nonlinear patterns associated with the predictor variables. We further discuss the summary of the final model together with visualization of the results.

The plots in Figure 3.2 illustrate the interaction between the effect of P-O Consistency and Time for unreduced forms (left panel) and reduced forms (right panel). Shades of blue indicate small pupil dilation; shades of green/yellow indicate large pupil dilation. The contour lines represent the pupil dilation values predicted by the model. These plots reveal that pupil dilation peaks around 800 ms for unreduced forms and 1000 ms for reduced forms. We observe modest P-O consistency effects around the peak dilation for both forms, but pupil dilation becomes greater as P-O consistency increases for unreduced forms, while pupil dilation becomes

Table 3.2: 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	6.919	13.084	0.529	0.596
Reduction:Reduced	21.819	4.001	5.453	< 0.0001
BaselinePupilSize	-0.025	0.003	-6.692	< 0.0001
WordDuration	76.117	20.132	3.781	< 0.0001
Smooth terms	**edf**	**Ref.df**	**F-value**	**p-value**
te(Time, P-O Consistency):Unreduced	8.741	8.882	57.832	< 0.0001
te(Time, P-O Consistency):Reduced	8.730	8.882	65.300	< 0.0001
s(Trail Index)	3.378	3.758	20.508	< 0.0001
s(GazeX, GazeY)	6.339	7.679	11.714	< 0.0001
s(Time, ParIDConsis)	1005.174	1793.000	1.905	< 0.0001
s(Time, ItemReduc)	720.878	2256.000	0.994	< 0.0001

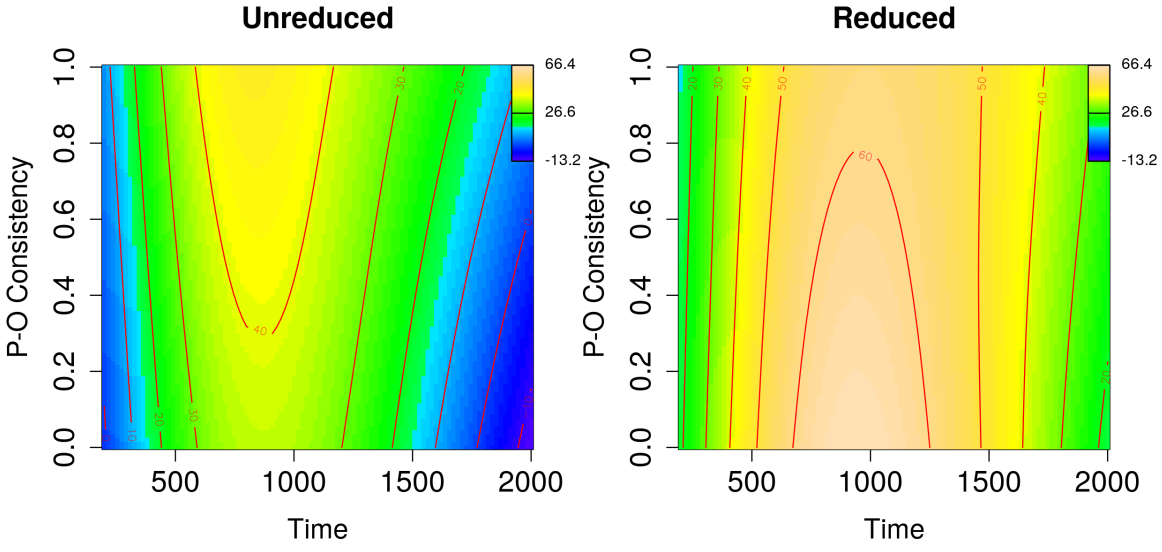

Figure 3.2: Contour plots of the interaction between the effect of P-O Consistency and Time for unreduced forms (left panel) and reduced forms (right panel). Shades of blue indicate small pupil dilation; shades of green/yellow indicate large pupil dilation. The contour lines represent the pupil dilation values predicted by the model.

greater as P-O consistency decreases for reduced forms.

Delayed naming task

Similar to the Go-NoGo task, Figure 3.3 displays the grand average of pupillary responses over time for reduced and unreduced forms from -1500 ms to 2500 ms in the delayed naming task. The trend of pupil dilation over time appears to be comparable between the two forms, but reduced forms demonstrate greater peak dilation. Error rates for naming responses were lower than 1% for both forms.

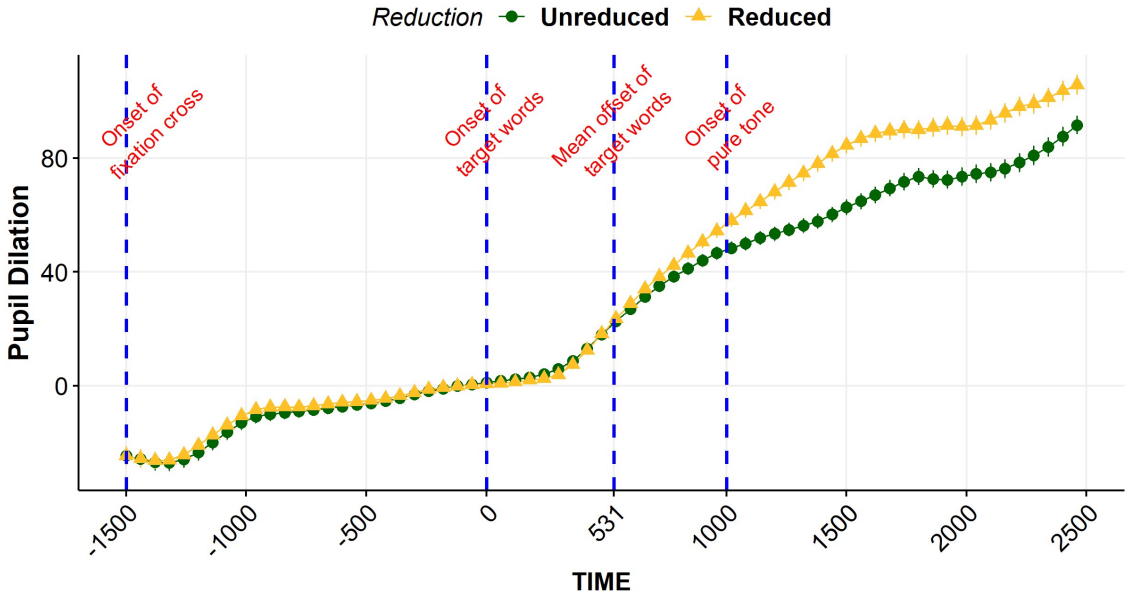

Figure 3.3: The grand average of pupillary responses over time for reduced and unreduced word forms in the delayed naming task. The vertical dot line at -1500 ms indicates the onset of the fixation cross, the line at 0 ms indicates the onset of stimuli, the line at 531 ms indicates the mean offset of stimuli, and the line at 1000 ms indicates the onset of pure tones.

For the delayed naming task, we chose the time window from 200 ms to 2500 ms post stimulus onset for our analyses. We used the same variables and procedures as earlier. For fixed effects, we eliminated Target Segment, Word Duration, and Z-transformed Number of Homophones. Logged Word Frequency and Baseline Pupil Size were refitted without a smooth function because their effects were linear. For random effects, we included ParIDConsis for Time and ItemReduc for Time. We also included an AR-1 correlation parameter at the value of 0.98 and fitted the model with the scaled-t family. The summary of our fi-

nal model is described in Table 3.3. The parametric coefficients indicate that the overall pupil dilation is greater for reduced forms than for unreduced forms ($t = 6.131, p < 0.0001$), the overall pupil dilation is smaller for larger baseline pupil sizes ($t = -6.624, p < 0.0001$), and that the overall pupil dilation becomes smaller as logged word frequency increases ($t = -2.401, p < 0.0001$). The smooth terms reveal the significance of non-linear patterns associated with the predictor variables. The summary of the final model, illustrated in Table 3.3, shows the parametric coefficients and approximate significance of smooth terms in the model. The parametric coefficients show that the overall pupil dilation is greater for reduced forms ($t = 2.381, p < 0.05$) and that the overall pupil dilation is smaller for larger baseline pupil size ($t = -565.960, p < 0.0001$). Similar to Go-NoGo task, the smooth terms indicate the significance of non-linear patterns associated with the predictor variables.

Table 3.3: 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	89.762	11.164	8.040	< 0.0001
Reduction:Reduced	14.065	2.294	6.131	< 0.001
BaselinePupilSize	-0.029	0.004	-6.624	< 0.0001
LoggedWordFrequency	-3.108	1.294	-2.401	< 0.05
Smooth terms	**edf**	**Ref.df**	**F-value**	**p-value**
te(Time, P-O Consistency):Unreduced	7.470	8.128	74.345	< 0.0001
te(Time, P-O Consistency):Reduced	8.337	8.707	98.498	< 0.0001
s(Trail Index)	3.194	3.633	38.177	< 0.0001
s(GazeX, GazeY)	8.720	8.983	30.459	< 0.0001
s(Time, ParIDConsis)	847.200	1704.000	1.372	< 0.0001
s(Time, ItemReduc)	586.034	2257.000	1.254	< 0.0001

Figure 3.4 illustrates contour plots of the interaction between the effect of P-O Consistency and Time for unreduced forms (left panel) and reduced forms (right panel). These plots reveal that the pupil dilates more greatly over time as we saw in the grand averages plot. We observe a modest effect of P-O consistency around

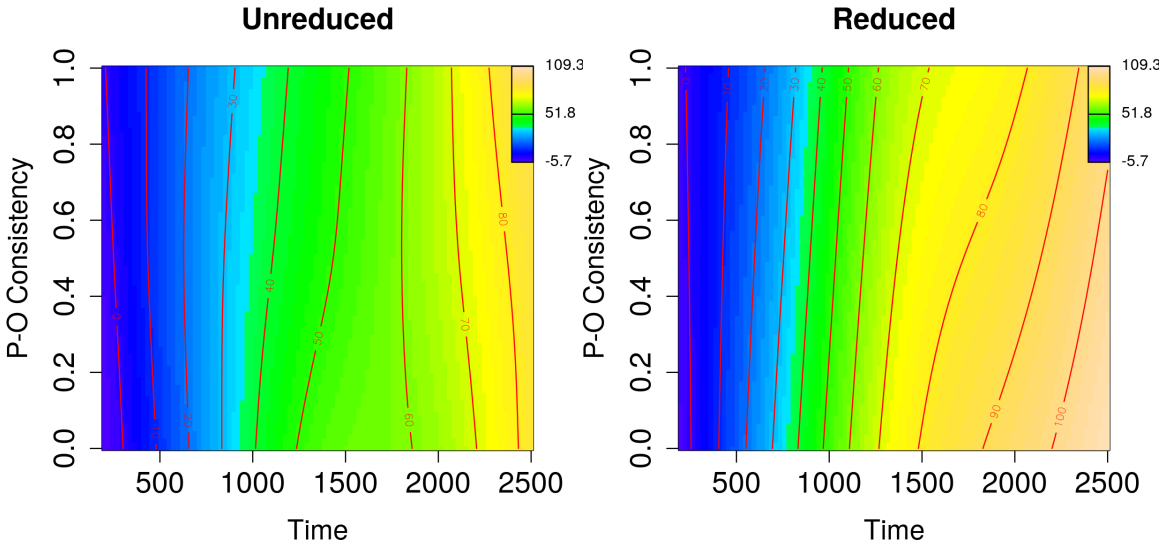

Figure 3.4: Contour plots of the interaction between the effect of P-O Consistency and Time for unreduced forms (left panel) and reduced forms (right panel). Shades of blue indicate small pupil dilation; shades of green/yellow indicate large pupil dilation. The contour lines represent the pupil dilation values predicted by the model.

1200 to 1500 ms for unreduced forms—pupil dilation values reach 50 earlier for low P-O consistency words than for high P-O consistency words. The effect becomes stronger in reduced forms in the time window from 1500 to 2500 ms—we observe steeper pupil dilation for low P-O consistency words than for high P-O consistency words, indicating an earlier and greater rise of dilation as P-O consistency decreases.

3.3 Discussion and conclusion

Using pupillometry, we examined how the P-O consistency effect interacts with phonetic reduction over time in Japanese. We used two types of tasks, Go-NoGo and delayed naming, together with the pupillometry, to compare the time course of the P-O consistency effect between reduced and unreduced forms of Japanese words. We predicted that if P-O consistency affects reduced and unreduced forms differently, we should observe an interaction between the effect of reduction and P-O consistency. Our findings are as follows: (1) reduced forms appeared to elicit larger dilation than unreduced counterparts in both tasks, (2) the P-O consistency

effect emerged in both tasks, and it influences reduced and unreduced forms differently, and (3) dilation becomes greater as the P-O consistency decreases. In what follows, we discuss each aspect of the results of both tasks considering our predictions and the results in previous studies.

First, we observed that reduced forms elicited greater pupil dilation than unreduced counterparts in both tasks, which is in line with the results in previous studies suggesting that processing is more effortful for reduced forms than for unreduced ones (Ernestus & Warner, 2011; Tucker, 2011; Warner & Tucker, 2011). Importantly, for the delayed naming task, the processing cost difference between the two forms varied depending on P-O consistency. That is, we observed a trend where the difference between reduced and unreduced forms is greater for inconsistent words (low consistency values) than for consistent words (high consistency values). In other words, the processing cost of reduced forms is prominent inconsistent words, but it attenuates or disappears in consistent words (Figure 3.2 and 3.4). This result is in line with the discussion of Racine et al. (2014); Ranbom & Connine (2007, 2011); Viebahn et al. (2018), and our results demonstrated that phonological and orthographic consistency plays an important role in the processing of reduced forms. We will further discuss an interaction between the processing cost of reduced forms and the effect of P-O consistency in Chapter 5. Second, for the Go-NoGo task, we found a modest effect of P-O consistency for both forms, suggesting that differing from the discussion of Mitterer & Reinisch (2015), the consistency effect also emerged with reduced forms. For the delayed naming task, while we observed the P-O consistency effect for both forms, the effect was particularly modest in unreduced forms, where pupil dilation values reached 50 earlier for low P-O consistency words than for high P-O consistency words within the time window from 1200 to 1500 ms. However, the effect became stronger in reduced forms and lasted longer, where steeper and greater pupil dilation was observed for low P-O consistency words than for high P-O consistency words in the time window from 1500 to 2500 ms. This suggests that the P-O consistency effect influenced the two forms differently and that the effect played an important role in the processing cost of reduced forms. A possible reason that we found the consistency effect for

reduced forms in both tasks, unlike the result in Mitterer & Reinisch (2015), is due to the difference in the degree of "conversational-likeness". While reduced forms are presented with informal sentences including discourse markers and contractions (more conversational) in their study, these forms are presented in isolation in our study (less conversational), suggesting that while the orthography plays a role in the recognition of reduced forms in isolation, the effect disappears when such forms are presented with more conversation-like context. This suggests that the way in which reduced forms are presented plays an important role in the effect of P-O consistency. Furthermore, in contrast to previous studies, we found the consistency effect in a naming task (Ventura et al., 2007, 2004; Ziegler et al., 2004). This could be because of the measurement utilized. Whereas these previous studies employed offline measures (i.e., reaction latency), we employed online a measure, specifically pupillometry. Our application of pupillometry to a naming task has revealed that the consistency effect emerges in the task.

Third, while the consistency effect was found in both forms in the Go-NoGo task, the direction of the effect differed between the two forms, suggesting that the P-O consistency effect influenced the two forms differently. While the processing cost rose as P-O consistency decreased for reduced forms, the cost rose as P-O consistency increased for unreduced forms. This direction of the effect found in unreduced forms was opposite to what has been found in previous studies (Hino et al., 2017; Ziegler & Ferrand, 1998; Ziegler et al., 2004). This discrepancy could be due to the fact that low P-O consistency words tend to have a higher number of phonological neighbours. In Japanese, a high phonological neighbourhood density facilitates the recognition of words (Yoneyama, 2002), meaning that the facilitatory effect of high phonological neighbourhood density is confounded with the inhibitory effect of low P-O consistency. That is, for unreduced forms, the facilitatory effect of high phonological neighbourhood density might have overwritten the inhibitory effect of low P-O consistency. This result is in line with the discussion of Rastle et al. (2011) where task difficulty could be driving force of the degree of activation of orthographic information, suggesting that P-O consistency might not have played an important role in the processing of unreduced forms because

they were easy to process, as indicated by a small pupil dilation. In contrast to the result of unreduced forms for the Go-NoGo task, the direction of the consistency effect was same for both forms and consistent with the previous studies in the delayed naming task. This could be because of the degree of processing cost required to perform the task. The processing cost required for unreduced forms in the delayed naming task was greater than for unreduced forms in the Go-NoGo task, as evidenced by the pupil dilation values (Figure 3.2 and 3.4). That is, orthographic information was activated and exerted even for the comprehension of unreduced words in the delayed naming task.

We also conducted a post-hoc analysis by taking a subset of data and found that the direction of P-O consistency effect changed depending on the number of phonological neighbors, suggesting that there is an interaction between the effect of P-O consistency and the number of phonological neighbors. Our further speculation for the direction of P-O consistency effect is that it might change depending on the type of orthography (logographic or alphabetic). Previous research has shown that in French (alphabetic), inconsistent words elicited increased negativity (N400) (Pattamadilok et al., 2009; Perre et al., 2011), but in Chinese (logographic), the effect is reversed such that consistent words elicited increased negativity (Chen et al., 2016). That is, the direction of P-O consistency effect could differ between logographic and alphabetic languages, and this seems to be apparent when employing an online measure, particularly ERPs.

In summary, we investigated how the P-O consistency effect interacts with phonetic reduction (casual versus careful speech) over time in Japanese. We found that P-O consistency influences reduced and unreduced forms differently in both tasks, and P-O consistency plays an important role in the processing cost of reduced forms. We also demonstrated that pupillometry can be used to investigate the effect of P-O consistency and reduced speech.

Chapter 4

The effect of phonological-orthographic consistency in the recognition of reduced speech for L2 speakers: Evidence from pupillometry

4.1 Introduction

Inconsistencies between the way in which words are pronounced and spelled have been shown to affect the recognition of spoken words (Ziegler & Ferrand, 1998). Inconsistent words are more difficult to process than consistent words. Inconsistent word are ones, where a single sound unit (e.g., rhyme) can be spelled in multiple ways, such as /-ip/ as in -eap or -eep in "leap" or "keep" and consistent words are ones, where a single sound unit can be spelled in only one way, such as /-ʌk/ as in -uck in "luck"). However, the situation is complicated by the fact that the realization of sound segments in spoken language is highly variable, especially in conversational speech (Warner & Tucker, 2011). Corpus research suggests that more than 60% of words are realized in variable forms and that 25% of these forms demonstrate segment deletion, as compared to a dictionary transcription of the word (Dilts, 2013; Johnson, 2004). In the third chapter, we found that reduced and unreduced word forms interact with the effect of inconsistency differently in the native speakers (L1) of Japanese. In the present study, we extend the investigation

to nonnative speakers (L2) and examine how the effect of inconsistency interacts with both reduced word forms and the proficiency of L2 speakers.

4.1.1 Orthographic effect in spoken word recognition

The influence of orthography in spoken word recognition has been observed by a number of studies, suggesting that orthographic representations are activated during the recognition of speech (Ziegler & Ferrand, 1998). Research has revealed that auditory rhyme detection is faster for similarly spelled word pairs (e.g., *tie - pie*) than for differently spelled pairs (e.g., *tie - rye*) (Seidenberg & Tanenhaus, 1979), and that auditory phoneme detection is slower for phonemes that have more possible spellings than phonemes that have fewer possible spellings (Frauenfelder et al., 1990). Additionally, primed lexical decisions are faster when primes and targets share both orthographic and phonological information than when they overlap only either phonologically or orthographically (Jakimik et al., 1985).

Orthographic consistency effect

Evidence for the influence of orthography in spoken word recognition has also been provided by studies on pronunciation and spelling inconsistency. The inconsistent relationship between the pronunciation and spelling of words (i.e., multiple ways to spell a single sound segment) has been shown to affect the rate at which listeners recognize spoken words (Stone et al., 1997; Ziegler et al., 1997). While this inconsistency does not occur with shallow orthographies, such as Finnish, it arises in deep orthographies, such as English and French (Frost, Katz & Bentin, 1987; Katz & Frost, 1992). For example, Ziegler & Ferrand (1998) measured the reaction time of auditory lexical decisions with consistent and inconsistent French words. The French word *stage* ('stage' in English) is consistent as the rhyme of the word /-aʒ/ has only one possible spelling -age, and the word *plomb* ('lead' in English) is inconsistent as the rhyme of the word /-õ/ has more than one possible spelling, such as -om, -ompt,-on. Their results revealed slower response times (62 milliseconds slower in mean reaction time) and higher error rates (13.1% higher) for inconsistent words than their consistent counterparts. Furthermore, Ziegler

et al. (2004) observed an effect of spelling probability within the consistency effect. For instance, the words *sign* and *wine* have the same rhyme /-aɪn/ but these two are spelled differently, meaning that both words are inconsistent. However, the -ine spelling is more common than the -ign spelling (e.g., *fine, nine, vine, dine*). That is, although both *sign* and *wine* show inconsistency, *wine* contains a more dominant spelling than *sign*. Ziegler et al. (2004) found the slowest response times and the highest error rates for inconsistent non-dominant spelling words, followed by the inconsistent dominate spelling words, and the fastest responses and lowest error rates for consistent words. They also compared the degree of the consistency effect as a function of the type of task using three tasks: auditory lexical decision, rhyme detection, and auditory naming task. The results indicated that the effect was strongest in the auditory lexical decision task, and the second strongest in the rhyme detection task, and the smallest (or null) in the auditory naming task (the effect reached significance only with by-subjects analysis in the naming task).

While it has been argued that the effect of orthography reflects strategic responses (i.e., the meta-phonological or meta-linguistic analysis) to the given task (Cutler, Treiman & van Ooijen, 2010), the consistency effect has also been demonstrated in a semantic task and even in a non-(meta)linguistic task. For the semantic task, Pattamadilok et al. (2009) employed ERP measurements with a Go-NoGo task. In a Go-NoGo task, a participant performs an action for particular stimuli (e.g., press a button, "Go") and inhibits the action for a different set of stimuli (e.g., do not press the button, "NoGo"). In the experiment of Pattamadilok et al. (2009), participants pressed a button for the name of a body part and they withheld their response for all others. In the NoGO trials, consistent, early inconsistent, and late inconsistent words were presented auditorily. As an example of early and late inconsistent French words, *rhume* ('cold' in English) is an early inconsistent word as the first two phonemes of the word /ʁy-/ have more than one possible spelling, and *noce* ('wedding' in English) is a late inconsistent word as the last two phonemes of the word /-ɔs/ has more than one possible spelling. The ERP measure revealed a time-locked orthographic consistency effect, indicating that early inconsistent words showed an early effect and late inconsistent words displayed

a late effect of inconsistency. The authors argued that since the effect emerged at a relatively early point in time, orthographic influences occur at the lexical access stage rather than decisional or postlexical stage.

Perre et al. (2011) also demonstrated the orthographic consistency effect using ERP measurements with a Go-NoGo non-(meta)linguistic task. In their experiment, participants pressed a button for white noise and did not press the button for French words. Although the participants did not have to make any linguistically oriented decisions on the French words, the consistency effect emerged, where inconsistent words indicated more negative ERPs than consistent words as early as 300 ms after the onset of spoken words. Importantly, Perre et al. (2009) compared three hypotheses on the nature of the consistency effect using ERPs and standardized low resolution electromagnetic tomography (sLORETA): (1) the online account hypothesizes that the bidirectional links between autonomous orthographic and phonological representations involuntarily co-activate each other (Dijkstra, Van Heuven & Grainger, 1998; McClelland & Rumelhart, 1981; Rumelhart & McClelland, 1982), (2) the restructuring account hypothesizes that orthographic information changes the phonological representation as literacy skills develop; that is, the acquisition of literacy leads to the restructuring of an individual's phonological representations, resulting in a single "phonological-orthographic" representation. (Muneaux & Ziegler, 2004; Taft, 2006, 2011; Ziegler et al., 2004, 2003), and (3) the hybrid account posits that both (1) and (2) occur simultaneously (Chen et al., 2016); as orthographic and phonological representations become more stable, the two representations are better integrated (Veivo & Järvikivi, 2013; Veivo et al., 2016). The findings of Perre et al. (2009) are as follows: the ERP measures indicated difference between consistent and inconsistent words as early as 330 ms after the onset of the target, and the standardized low resolution electromagnetic tomography (sLORETA) indicated a differential activation at 350 ms between the two words in a left temporo-parietal area, shown to respond to phonological processing (Duffau, 2008), but no activation in the occipito-temporal brain area, shown to play a role in the processing of the visual word-form information (Cohen & Dehaene, 2004). This result is in line with the restructuring hypothesis, and Perre

et al. (2011) concluded that orthographic information contaminates phonological representations.

Reduced speech and orthographic effects

Almost all of the previous research has examined the consistency effect using carefully pronounced words. Spoken word forms are, however, highly variable, particularly in casual everyday speech. This variability is often the result of phonetic reduction, realized as shortening, deletion, and/or incomplete articulation of segments (Ernestus & Warner, 2011; Greenberg, 1999; Warner & Tucker, 2011). For example, *yesterday* may be realized as [jɛʃeɪ] (Tucker, 2007) and 原因 *genin* 'reason' may be realized as [gẽːin] in Japanese (in the second chapter). These examples suggest that the actual pronunciation (phonetic form) of words could be substantially inconsistent with the spelling, meaning that the inconsistency between the way *yesterday* is pronounced and spelled should be greater for the casual pronunciation [jɛʃeɪ] than for the careful pronunciation /jɛstəˇdeɪ/. This additional inconsistency created by reduced forms is ubiquitous, as research shows that more than 60% of words in the Buckeye Corpus of Conversational Speech (Pitt et al., 2007) were realized in variable forms, and 25% of these forms demonstrated segment deletion as compared to a dictionary transcription of the word (Dilts, 2013; Johnson, 2004). Interestingly, these reduced forms are recognized less efficiently than canonical/unreduced counterparts despite the fact that reduced forms occur more frequently than canonical/unreduced ones (e.g., Arai et al., 2007; Ernestus et al., 2002; Tucker, 2007, 2011; van de Ven et al., 2011), unless supportive discourse contexts are provided (Brouwer, Mitterer & Huettig, 2013). Recent research has argued that the difficulty in processing reduced variants disappears when the frequency of these variants is accounted for (Bürki, Viebahn, Racine, Mabut & Spinelli, 2018) and 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 the orthographic form of words (Bürki et al., 2012; Charoy & Samuel, 2019; Racine et al., 2014; Ranbom & Connine, 2007, 2011; Viebahn et al., 2018).

Mitterer & Reinisch (2015) extended the investigation of the orthographic effect on reduced variants by looking at conversational speech, where reduced forms were presented with conversational-speech like context. They compared the processing cost of the 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 perception of spoken words, the deletion of the orthographically uncoded segment should have a smaller impact on the processing cost than the deletion of the orthographically coded segment. In order to test the hypothesis, they conducted a visual world eye-tracking experiment and pronunciation judgment task. In the eye-tracking experiment, participants saw pictures of three objects and an empty quadrant on a screen. They then heard a sentence including a target word. Participants clicked on one of the pictures to identify the word mentioned in the sentence. If none of the objects on the screen were mentioned, they clicked on the empty quadrant. The target words were recorded by native female speakers of the two languages, and the researchers instructed the native speakers to produce these words casually. These words were then manipulated to create a reduced form of the words. Importantly, the carrier sentences included discourse markers (e.g., *like*) and contractions (you've instead of you have) to simulate an informal conversational speech style. The results indicated that the processing cost caused by the reduction of orthographically coded and uncoded segments did not differ, suggesting that orthographic coding did not affect the processing cost. In the pronunciation judgement task, native speakers of the two languages recorded the same words, but the researchers instructed them to produce these words clearly. The researchers then manipulated these words to create reduced forms. Instead of using full sentences, short phrases were used to convey the target words without discourse markers and contractions. In this task, participants saw a printed target word on a computer screen and heard a short phrase containing the target word. They then rated how good the pronunciation of the target word was on a scale from 1 (very bad) to 7 (very good). In contrast to the findings for conversational-style speech, Mitterer & Reinisch (2015) found the effect of the processing cost caused by reduction, the orthographic coding of

segments, and the interaction between the two, suggesting that orthographic coding of the target words affected the recognition of spoken words. That is, deletion lowered the ratings of pronunciation and more so for the orthographically coded segment /h/ than uncoded segment /ʔ/. As a result, the researchers concluded that orthography does not play an important role in perceiving 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 & Reinisch (2015) also noted a few issues to be addressed in future work. One of the relevant issues for the present research is to describe how a task type and speech style interact to activate orthography, since the contrast between their two experiments confounds speech styles with task types. They did not find the effect of orthography with an implicit task (visual world paradigm) using casually uttered words but found the effect with an explicit task (pronunciation judgement) using carefully pronounced words.

Orthographic effects in L2 speakers

The research review thus far suggests that orthography influences the recognition of spoken words, and the pronunciation and spelling consistency also matters, but the orthographic effects are likely to disappear in conversational speech. These findings are, however, based on the assumption that phonological representations are acquired by L1 speakers prior to orthography. However, many L2 speakers acquire phonological and orthographic information simultaneously or their acquisition of orthographic representation may even precede that of phonological representations. Previous research shows that the phoneme and grapheme relationships established during the process of learning novel words influences the perception of L2 sounds (Escudero, Hayes-harb & Mitterer, 2008; Escudero & Wanrooij, 2010; Hayes-Harb, Brown & Smith, 2018). Furthermore, using a masked cross-modal priming experiment, Veivo & Järvikivi (2013) investigated the role of orthography in the perception of spoken words with L2 French speakers of L1 Finnish. They employed L2 visual primes (French) for L2 auditory targets (French) in three condi-

tions: repetition prime (same as the target), pseudohomophone prime (non-word that could be pronounced like the target), and non-word control prime (no overlap with the target). They found that both repetition and pseudohomophone primes facilitated the recognition of the targets, but the effect of the pseudohomophone primes was smaller. In addition, they employed L1 visual primes (Finnish) for L2 auditory targets (French) in the following conditions: orthographic onset overlap prime (Finnish word that orthographically overlaps in three first letters with the target but is semantically unrelated), Finnish pseudohomophone prime (non-word that could be pronounced like the target), and unrelated Finnish prime (no semantic, phonological, or orthographic overlap with the target). Their results indicated that the orthographic overlap primes elicited a facilitatory effect with highly proficient participants, but this effect disappeared with highly familiar target words. Moreover, for the lower intermediate proficiency participants, the Finnish pseudohomophone primes elicited a facilitatory effect but the orthographic overlap primes did not. As a result, they concluded that L1 orthography plays a role in the perception of spoken words in an L2, but the effect of L1 orthography depends on the proficiency of L2 speakers because the lower intermediate learners have not yet fully established lexical representations of L2 words.

Additional research by Veivo et al. (2016) employed a visual world eye-tracking experiment. L2 French speakers of L1 Finnish and L1 French speakers matched spoken targets (French) with printed words (French) by clicking on a target word on a computer screen. The target word appeared on the screen together with competitors varying in length of word initial phonological overlap and word initial orthographic overlap. For example, for the target word (*base* /baz/ 'base'), the higher degree phonological overlap competitor was (*bague* /bag/ 'ring'), and the lower degree phonological competitor was (*bain* /bɛ̃/ 'bath'), and for the target word (*mince* /mɛ̃s/ 'slim'), the higher degree orthographic overlap competitor was (*mite* /mit/ 'moth'), and the lower degree orthographic competitor was (*mythe* /mit/ 'myth'). The results revealed an effect of phonological overlap in L2 French speakers. The longer phonological overlap competitors affected the mapping process more than the shorter phonological overlap competitors. Additionally, although

there was no main effect of the degree of orthographic overlap (short vs. long), the results showed that the effect of orthographic overlap over time was dependent on L2 proficiency. The longer orthographic overlap competitors affected the mapping process more than the shorter orthographic overlap competitors in the time-window of 400 to 700 ms, and this effect appeared only with highly proficient learners. As a result, they concluded that orthography has a role to play along with L2 proficiency in the visual search for printed words in late learners of L2 French. Similar to Veivo & Järvikivi (2013), the authors argued that this is because phonological and orthographic knowledge in low proficiency learners are not yet fully established; therefore, these two types of knowledge do not interact effectively. In other words, the two types of knowledge are activated somewhat separately, but once learners become highly proficient, these two types of knowledge are integrated and interact efficiently. This suggests that low proficiency learners co-activate orthographic and phonological representations, as the on-line hypothesis suggests, and high proficiency learners integrate orthographic information into phonological representation, as the restructuring hypothesis argues, as discussed earlier in the section of orthographic consistency effect. Finally, Veivo et al. (2016) also argued that there might be a qualitative difference between L1 and L2 speakers' lexical knowledge due to the early introduction of orthography in L2 learning, which builds a strong orthographic component in the learners' lexical knowledge. Similar results were also found in the effect of L1 orthography in L2 spoken word recognition (Veivo, Porretta, Hyönä & Järvikivi, 2018).

Orthographic effects in logographic languages

While all of the studies discussed above are based on alphabetic languages, languages that utilize logographic scripts, such as Chinese and Japanese, also show the effect of consistency. In such languages however, the pronunciation and spelling mappings are not straightforward because in the Japanese writing system for example, a logographic character cannot be decomposed into smaller orthographic elements that correspond to phonological units which are smaller than the whole character. In other words, Japanese logographs do not reliably correspond to any

sub-lexical phonological unit, such as phonemes or morae (Fushimi et al., 1999; Wydell, 1998; Wydell et al., 1993). Japanese orthography consists of two types of scripts, phonographic (or syllabary) *Kana* and morphographic (or logographic) *Kanji*. Kana comprises two types of syllabaries, Hiragana and Katanana, each of which consists of 46 base scripts—5 vowels, 40 consonant and vowel combinations, and 1 single consonant, and each syllabary represents a single mora of a spoken word (e.g., Fushimi et al., 1999; Wydell et al., 1993). In modern Japanese, kanji scripts are usually used to write nouns and stems of both adjectives and verbs. While hiragana scripts are generally used to complement grammatical elements of words (e.g., inflections of verbs and adjectives, grammatical cases, and phonetic complements to disambiguate pronunciation), katakana scripts are mostly used to represent loanwords. Additionally, the majority of kanji characters have two types of pronunciations, ON and KUN, determined by the specific combination of the two (or more) kanji characters, and many kanji characters also have multiple ON pronunciations and occasionally even more than one KUN pronunciation (Fushimi et al., 1999; Wydell et al., 1993). Finally, although kanji words can contain many kanji characters, many kanji words comprise two kanji characters (Fushimi et al., 1999).

While logographic characters cannot be decomposed into smaller orthographic elements that correspond to sub-lexical phonological units, they can be decomposed into smaller sub-character components, one of which is the phonetic radical which often provides a clue to the pronunciation of the whole character. The reliability of the phonetic radicals has been utilized to define the pronunciation and orthographic consistency for logographic languages. Using ERPs and standardized low-resolution electromagnetic tomography (sLORETA), Chen et al. (2016) investigated the effect of consistency as measured by homophone density and orthographic consistency in Chinese. Homophone density represents the number of characters that share the same pronunciation, and orthographic consistency refers to the degree to which a set of homophones can be categorized into subgroups on the basis of their phonetic radicals. The ERP measures revealed both effects of the orthographic consistency and homophone density, and the sLORETA indi-

cated the orthographic effect both in the phonological processing region (frontal and temporal-parietal region) and in the orthographic processing region (posterior visual cortex region), which is in line with the hybrid account, where the activation of orthographic representations and the restructuring of phonological-orthographic representations occurs simultaneously. Similar orthographic effects were also found both in L1 (Qu & Damian, 2017) and L2 Chinese speakers (Qu et al., 2018).

Phonological-orthographic consistency effect in Japanese

While several studies have investigated the orthography to phonology (O-P) consistency effect in Japanese (Fushimi et al., 1999; Hino, Miyamura & Lupker, 2011; Saito, Masuda & Kawakami, 1998; Wydell, Butterworth & Patterson, 1995; Wydell et al., 1993), research on the phonology to orthography (P-O) consistency effect is very limited. One of the experiments in Hino et al. (2017) investigated the P-O consistency effect using the P-O consistency index. Similar to the O-P consistency index in Hino et al. (2011), they calculated the P-O consistency index based on the frequency of phonological and orthographic neighbours of the target words, which is similar to the approach taken by Fushimi et al. (1999) that follows the procedures employed in Jared et al. (1990). They first identified the phonological neighbours of a 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 as 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. For example, if the target word is 現在 (/ge-n-za-i/, 'current'), the phonological neighbours of the target word are 犯罪 (/ha-n-za-i/, 'crime'), 限界 (/ge-n-ka-i/, 'limit'), 現代 (/ge-n-da-i/, 'present day') and 存在 (/so-n-za-i/, 'existence'), as all of these words differ by a single mora from the target word. [2] Of these words, 現代 (/ge-n-da-i/, 'present day') and 存在

[2]Hyphens indicate mora boundaries.

(/so-n-za-i/, 'existence'), are orthographic friends as they differ by a single character from the target word, and 犯罪 (/ha-n-za-i/, 'crime') and 限界 (/ge-n-ka-i/, 'limit') are orthographic enemies, as more than one character is different from the target word. Following this classification, the frequencies of the target word and orthographic friends are summed and divided by the sum of the frequencies of the target word, orthographic friends, and orthographic enemies. The resulting 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 the effect of P-O consistency, where response latencies for low consistency words were slower than for high consistency words.

Similar orthographic effects were also found in L2 Japanese speakers. Using a visual world eye-tracking experiment, Mitsugi (2018) investigated the effect of orthography and its time-course in the recognition of L1 and L2 spoken words with L2 Japanese speakers of L1 English. Similar to Veivo et al. (2016), L2 Japanese speakers of L1 English and L1 Japanese speakers matched spoken Japanese target words with printed Japanese Hiragana words by clicking on a target word on a computer screen. The target word appeared on the screen together with one phonological competitor that shares the initial mora with the target word and two distractors. For example, when さかな (/sa.ka.na/, 'fish') was the target word, さくら (/sa.kɯ.ɾa/, 'a cherry blossom') was the phonological competitor, and むかし (/mɯ.ka.ɕi/, 'long time ago') and つもり (/t͡sɯ.mo.ɾi/, 'intention') were distractors. Mitsugi (2018) found the orthographic effect for L1 speakers, and for L2 speakers, the effect was dependent of L2 proficiency, where higher proficiency speakers looked to the target words (vs. the competitors) more compared to lower proficiency speakers. Yet, she also reported that the task employed in the experiment was limited to identify the recognition of the precise trigger of the effect of orthography.

In summary, previous research has suggested that there are effects of orthography in spoken word recognition in both alphabetic and logographic languages in both L1 and L2 speakers. In the present study, we further investigate the orthographic effect by focusing on the effect of P-O consistency in L2 Japanese speakers. We further investigate (1) whether the claims made by Mitterer & Reinisch (2015) extend to Japanese, which employs logographic scripts and also (2) whether it extends to L2 speakers. Specifically, we investigate how the P-O consistency effect interacts with both reduced word forms and the proficiency of L2 speakers. First, if the orthographic effects do not play an role in the recognition of reduced speech, as Mitterer & Reinisch (2015) argue, we should observe the P-O consistency effect in unreduced forms but not in reduced forms. Or possibly, the consistency effect influences the two types of word forms differently, as seen in the results of L1 speakers in the third chapter, suggesting that the surface word form matters for the effect. Second, if phonological and orthographic representations in low proficiency learners do not interact efficiently due to the fact that these representations are not yet fully established as Veivo & Järvikivi (2013) and Veivo et al. (2016) suggest, we should observe a stronger consistency effect in high proficiency learners as opposed to low proficiency learners. Or, it is possible that since L2 learners have a strong orthographic component in their lexical knowledge due to the early learning of orthography, they would show a strong P-O consistency effect (Qu et al., 2018), possibly due to the reliance on orthographic information to compensate for the lack of stable phonological representations. This reliance on orthography may be particularly salient in low proficiency learners.

4.2 The current study

In the present study, we designed our experiments in such a way that we can address these issues by comparing the time course of the P-O consistency effect between reduced and unreduced words as indicated by pupil dilation. The pupil has been shown to respond to physiological arousal during cognitive tasks, and pupillometry, the measurement of pupil dilation, has been utilized as an index of

cognitive effort (Kahneman & Beatty, 1966). Pupillometry has been applied to a variety of psycholinguistic studies particularly recently as it offers a reliable method to measure cognitive effort, attention, and affect imposed by different variables in speech processing (Laeng et al., 2012) and it is largely free from the effect of task-specific strategies (Goldinger & Papesh, 2012). For our experiments, pupillometry is particularly beneficial because it reflects the extent to which our variables of interest impact the amount of cognitive effort over time in the absence of voluntary processes (Papesh & Goldinger, 2012, 2015). The activities within the locus coeruleus, which is part of the noradrenergic system, are correlated with two modes of pupillary responses: phasic and tonic. While the phasic pupillary response is time locked to task-related events and stimuli, the tonic (or baseline) response is slow-changing and is related to the state of arousal or vigilance (Laeng et al., 2012; Papesh & Goldinger, 2015). The phasic pupillary response has been utilized in a variety of psycholinguistic studies together with various types of tasks. With a lexical decision task, Kuchinke et al. (2007) revealed a frequency effect, where peak dilation is greater for low frequency words than for high frequency words. Goldinger & Papesh (2012) also found a frequency effect with a naming task, and Geller et al. (2016) observed an orthographic inhibitory effect, with lexical decisions in the masked priming. In these studies, stimuli were presented visually. Klingner, Tversky & Hanrahan (2011) shows that pupillary response is larger for auditory stimuli than for visual stimuli (see Zekveld, Koelewijn & Kramer (2018) for an overview of pupillometry research with auditory stimuli). Kramer, Lorens, Coninx, Zekveld, Piotrowska & Skarzynski (2013); Zekveld, Heslenfeld, Johnsrude, Versfeld & Kramer (2014); Zekveld et al. (2010) have revealed the relationship between speech intelligibility and listening effort using a speech reception threshold test, where the peak dilation, peak latency, and mean pupil dilation systematically increase with decreasing speech intelligibility, and Porretta & Tucker (2019) have also demonstrated the relationship between speech intelligibility and listening effort using a listen-and-repeat task, where listening effort increases as intelligibility decreases due to foreign accentedness (See Winn et al. (2018) for an overview of pupillometry research on listening effort) Lastly,

pupil dilation in response to both linguistic and extra-linguistic variables also reflect individual differences. Lõo, Järvikivi, Tomaschek, Tucker & Baayen (2018) has revealed that pupillary response during a word naming task reflects individual differences in the frequency effect, and Hubert & Järvikivi (2019) revealed that language comprehension is influenced by an individual's value or belief system.

Taken all together, we apply pupillometry to compare the time course of the P-O consistency effect with reduced and unreduced Japanese words. If P-O consistency affects reduced and unreduced forms differently, we should observe an interaction between the effect of reduction and P-O consistency. If the effect of P-O consistency varies depending on L2 proficiency, we should observe an interaction between the effect of P-O consistency and L2 proficiency. We chose a Go-NoGo and delayed naming task for our experiment as these tasks fit the pupillometry study and our research questions.

4.3 Method

We utilized a Go-NoGo task, where participants only respond to a particular set of stimuli (Perre et al., 2011). In our task, participants responded to non-target items (a pure tone) by pressing a button on a pad, and they do not respond to the target items (i.e., they were passively listening to the items). In this way, the target trials were free of artifacts resulting from motor movements invoked by responses that possibly influence pupil dilation, and the participants did not need to make any linguistically derived decisions on the the target items. We also utilized a delayed naming task, where participants hear a target word and wait for a response signal. After the signal, they repeat what they have heard (Forster & Chambers, 1973). In this task, participants hear a Japanese word and wait until they hear a pure tone. They then repeat what they have heard. Since the Go-NoGo task requires only passive listening for target items, we expect overall pupillary responses to be greater for the delayed naming task than for the Go-NoGo task.

4.3.1 Participants

Twenty-six native speakers of English (female, n = 16) speaking Japanese as an additional language were recruited at Nagoya University in Japan, 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. All participants performed both tasks. Task presentation order was split so that half performed the Go-NoGo task first and the other half did the delayed naming task first.

4.3.2 Materials

We chose 226 four-mora and two-logograph words (lists of the words are available on Education and Research Archive: https://doi.org/10.7939/r3-60xn-qd28) and used the Balanced Corpus of Contemporary Written Japanese (BCCSJ) (Maekawa et al., 2014) to collect information needed to calculate the P-O consistency index for each target word (frequency and number of phonological and orthographic neighbours). We selected BCCSJ because it holds one of the largest Japanese datasets and encompasses a more recent and wider range of styles and genres than the database (National Language Research Institute, 1970, 1993) used in Hino et al. (2017, 2011). Following Fushimi et al. (1999); Hino et al. (2017, 2011), phonological neighbours were defined as words that differ by a single mora from the target word, and orthographic neighbours were defined as words that differ by a single character from the target word. All of the target words contain a word-medial nasal and/or voiced stop because previous research has demonstrated that both types of consonants show various forms of reduction as discussed in the second chapter of this dissertation. For example, in the case of the word-medial nasal in /tenisu/ 'tennis', the nasal and following /i/ are deleted, and the /e/ is both nasalized and lengthened; the word is then realized as [tẽːsɯ] (Arai, 1999). For word-medial voiced stops, articulation of the word-medial voiced stop in /dɑigɑkɯ/ 'university' is approximated due to the lack of full oral closure and realized as [daiɣɑkɯ], and in the extreme case, the consonant is deleted and the realization of the word becomes [daiakɯ] (Arai, 1999). Our stimuli were recorded in both reduced and unreduced

forms by a female native Japanese speaker (452 items in total). We instructed the speaker to produce the words clearly (careful speech like) for unreduced forms and casually (spontaneous speech like) for reduced forms. The speaker produced multiple tokens of both forms, and we selected the most natural sounding tokens as stimuli. We then normalized the amplitude of the words for presentation purposes. Table 1 illustrates the acoustic properties of both reduced and unreduced forms. These are the same stimuli as used for the study of native speakers in Chapter 3. We defined the intensity difference as the difference between the minimum intensity of the target segment to the averaged maximum intensity of surrounding segments (Warner & Tucker, 2011). Overall, reduced forms have shorter duration, faster speech rate, lower mean pitch, and smaller intensity difference. Using two sample t-tests, we found that the differences between the two forms reached significance for all properties except the intensity difference. Importantly, the difference between the two word forms in our study is that the target segment is produced as reduced or unreduced as defined in Tucker (2011), rather than absence or presence of the segment. Previous studies on the effect of orthography in the recognition of reduced forms treated reduction processes as a phonological phenomena, such as the absence or presence of word-medial schwa in French (Bürki et al., 2012, 2018). While this is true for some cases, many reduction processes appear to be phonetic and gradient (Bürki, Fougeron, Gendrot & Frauenfelder, 2011; Ernestus & Warner, 2011; Warner & Tucker, 2011). For example, sound segments may be articulated rapidly and weakly, resulting in the shorter segments (reduction rather than deletion) as discussed in the second chapter and/or may contain strongly coarticulated segments, indicating phonetic properties that are not usually found in carefully pronounced words (Arai, 1999).

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 manipulated the ratio between the target and non-target trials as 70% and

SegmentType	Reduction	WordDuration	SegmentDuration	SpeechRate	MeanWordPitch	IntDifference
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	$p < 0.001$	$p < 0.001$	$p < 0.001$	$p < 0.001$	$p = 0.178$

Table 4.1: Mean acoustic values of stimuli in reduced and unreduced forms. Both word and segment durations are represented in seconds. Speech rate was measured by the number of vowels per second. p *values indicate the probability that the difference between the two forms were significantly different from 0.*

30%. The target words were counterbalanced across both reduction and task, so that none of the participants heard the same word twice.

4.3.3 Apparatus and procedure

We designed and controlled the experiment using SR Research Experiment Builder software. Participants' right eye was tracked by a 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 x 768 resolution computer screen to present a fixation cross. 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. For the delayed naming task, we used a head-mounted Countryman E6 microphone and Korg digital recorder to record participants' naming responses.

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

2012). As a result, participants' task was to identify a pure tone; therefore, participants did not make any linguistically derived decision on the target stimuli, allowing us to eliminate the possibility that P-O consistency is exclusively driven by linguistically motivated decision-making (Perre et al., 2011).

In the delayed naming task, participants looked at a fixation cross presented at the centre 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 500ms-long pure tone. They then repeated what they had heard. The fixation cross disappeared 2000 ms after the onset of Japanese words. 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 for 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, participants answered a Language Experience And Proficiency Questionnaire (LEAP-Q) (Marian, Blumenfeld & Kaushanskaya, 2007) to provide their demographic information, 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 behavioural performance data, particularly for L2 speakers, suggesting that self-reported proficiency reflects approximations of overall language skill (see Kaushanskaya, Blumenfeld & Marian (2019) for an overview of LEAP-Q).

4.3.4 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 and their artifacts, we cleaned the data by removing 50 samples (100 ms time window) before and after the blinks using Jacolien van Rij's *removeBlinks* function. 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. Three participants were excluded from our data in both tasks due to excessive blinks and their artifacts (more than 50% of trials contained more than 30% of eye-blinks and their artifacts). 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 location 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. Additional trials were excluded when the peak latency was shorter than 400 ms, the peak dilation was smaller than 0 or bigger than 400. In total, we excluded 23% of data in the Go-NoGo task and 13.7% of data in the delayed naming task. The data is available on Education and Research Archive (https://doi.org/10.7939/r3-60xn-qd28).

4.3.5 Variables of interest

Our dependent variable of interests 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 subtract 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). Our primary variables of interest for independent variables were P-O Consistency Index (ranging from 0 - 1), Self-Rated Overall L2 Proficiency (Basic, Intermediate, or Advanced), Reduction (Reduced

or Unreduced form), and Time (in milliseconds). The P-O consistency index was calculated based on the frequency of phonological and orthographic neighbours of a target word (Hino et al., 2017, 2011). After identifying all phonological neighbours of a target word, the phonological neighbours were classified into two types: orthographic friends and orthographic enemies (Jared et al., 1990; Ziegler et al., 2003). 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. Following this classification, the frequencies of the target word and orthographic friends are summed and divided by the sum of the frequencies of the target word, orthographic friends, and orthographic enemies. The P-O consistency index ranges from 0 to 1, with 0 indicating low consistency and 1 indicating high consistency (see Hino et al. (2017, 2011), for an overview of the P-O consistency index).

The self-rated overall L2 proficiency was collected via LEAP-Q. The questionnaire includes four proficiency levels: Basic, Intermediate, Advanced or Superior. We, however, only had one participant self-report as superior; therefore, the participant was included in the advanced group. Reduction is a binary category and represents whether the participant heard a reduced and unreduced stimulus. Finally, we were also interested in the time course of processing; therefore, Time (in milliseconds) was included as a covariate. In order to investigate the interaction between Self-Rated L2 Proficiency and Reduction, a new factor with six levels was created (ProficiencyReduction): Advanced.Unreduced, Basic.Unreduced, Intermediate.Unreduced, Advanced.Reduced, Basic.Reduced, and Intermediate.Reduced (see van Rij (2015); Wieling (2018); Wieling et al. (2016), for an overview of how to handle interactions in GAMM).

Additionally, we included control variables: 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 (from 6 to 150), Word Duration (in milliseconds), Target Word Frequency (log-transformed), Number of Phonological Neighbours of Target Word (log-transformed), and Number of Homophones of Target Word (Z-transformed). The baseline pupil size was included to account for the extent to

which the pupil can dilate, as it depends on the baseline pupil size (i.e., a larger baseline size limits the extent of dilation). Pupil Gaze Coordinate accounts for the possible change in pupil size caused by different gaze locations on the screen (Wang, 2011). Trial index was included to control for the effect of trial order, and Word Duration, Logged Target Word Frequency, Logged Number of Phonological Neighbours 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 Neighbours was negatively correlated with P-O Consistency Index ($r = -0.66$). That is, high consistency words tends to have a small number of phonological neighbours. Therefore, these two variables would not be included in the same model.

4.3.6 Statistical considerations

We applied Generalized Additive Mixed Modeling (GAMM) to our pupil dilation data for three reasons (Hastie & Tibshirani, 1990; Wood, 2006). First, GAMM allows us to model non-linear relationships, as well as linear relationships, between a response variable and predictor variables (Sóskuthy, 2017; 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 nonlinear interactions of continuous variables. Third, GAMM allowed us to control for serial dependency in time-series data, namely, autocorrelation (Baayen et al., 2017). It considers the correlation between an observed value at time point t and an observed value at time point $t+i$ ($i > 1$) in a time series (see Baayen et al. (2017) and Wood (2017) for an overview of autocorrelation in GAMM). Because of this advantageous functionality, GAMM has been utilized not only to model pupillometic data (Lõo et al., 2018; Mukai, Järvikivi & Tucker, 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), format trajectory data, the time-course of formant frequencies in speech (Sóskuthy, 2017), visual

world eye-tracking data (Porretta, Tucker & Järvikivi, 2016; Veivo et al., 2016), and event-related potential data (Kryuchkova, Tucker, Wurm & Baayen, 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 (2015); Wieling (2018); Wieling et al. (2016).

We employed a backwards stepwise elimination 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 input variables that were justified by the comparison ($p<.05$). Inclusion of interactions was also assessed by the fREML score comparison. For random effects, we employed a forward fitting model selection procedure to determine the optimal random-effects structure (Matuschek et al., 2017).

4.3.7 Results and Discussion

In the sections that follows, we discuss the statistical analysis and result of each task: first the Go-NoGo task and second the Delayed naming task.

Go-NoGo task

We inspected the aggregated raw pupil dilation data prior to fitting models. Figure 4.1 illustrates the grand average of pupillary responses over time for reduced and unreduced forms from -1500 ms to 2000 ms in the Go-NoGo task. The trend of pupil dilation over time appears to be comparable between the two forms despite the fact that the reduced form demonstrates slightly greater peak dilation.

We chose the time window from 200ms to 2000 ms post stimulus onset for the Go-NoGo task for our analysis, in that 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 In-

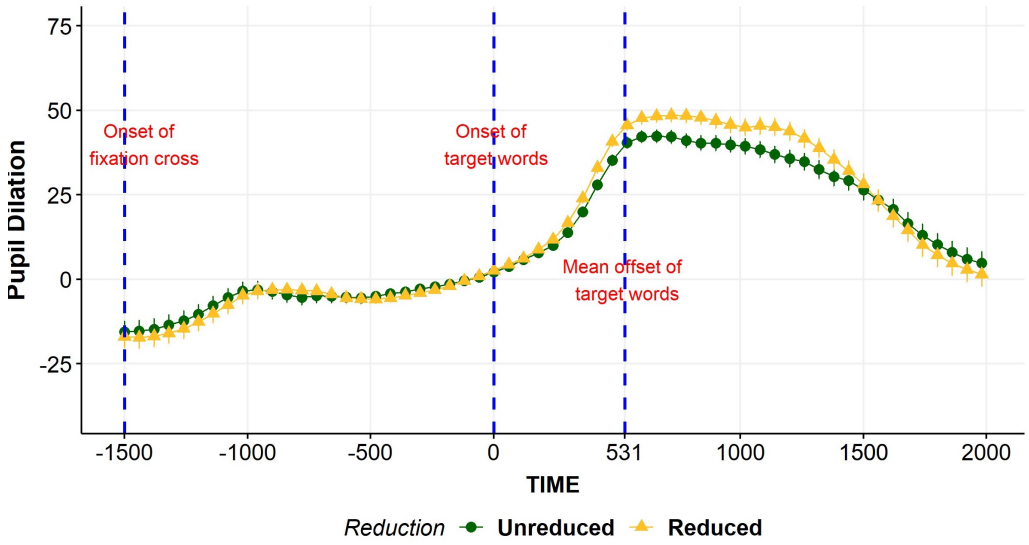

Figure 4.1: The grand average of pupillary responses over time for reduced and unreduced word forms in the Go-NoGo task. The vertical dot line at -1500 ms indicates the onset of the fixation cross, the line at 0 ms indicates the onset of stimuli and the line at 531 ms indicates the mean offset of stimuli.

dex, ProficiencyReduction, and Time (independent variables), with Baseline Pupil Size, Pupil Gaze Coordinate X and Y, Trial Index, Word Duration, Logged Word Frequency, Logged Number of Phonological Neighbours, and Standardized Number of Homophones (control variables) for each task. Furthermore, using a tensor product, a few types of interaction were included: three-way interaction for Time, P-O Consistency Index, and ProficiencyReduction, as well as two-way interaction for 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, P-O Consistency Index and ProficiencyReduction allows us to examine the effect of reduction, P-O Consistency Index and L2 proficiency, as well as the interactions between 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). For 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. For fixed effects, we excluded Word Duration, Logged Target Word Frequency, Logged Number of Phonological Neighbours, and Standardized Number of Homophones from the model during model fitting. After verifying the number of basis functions for the predictor variables and interactions using the *gam.check* function (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), we included an AR-1 correlation parameter p = 0.982 to address autocorrelation and refitted the model with *scaled-t* family in order for residuals to be normally distributed (Meulman et al., 2015; van Rij et al., 2019; Wieling, 2018).

Table 4.2 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 parametric coefficients indicate that the overall pupil dilation is greater for Intermediate.Unreduced ($t = 2.337, p < 0.05$), Advanced.Reduced ($t = 2.600, p < 0.01$), Basic.Reduced ($t = 2.405, p < 0.05$), and Intermediate.Reduced ($t = 2.405, p < 0.05$) than Advanced.Unreduced (the reference level). The smooth terms reveal the significance of non-linear patterns associated with the input variables. We further discuss the summary of the final model together with visualization of the results. The contour plots in Figure 4.2 illustrate the interaction between the P-O Consistency Index, L2 Proficiency, and Reduction over Time for the different combinations of participant group and word form: Basic proficiency participants with reduced forms (Top Left panel), Basic proficiency participants with unreduced forms (Top Right panel), Intermediate proficiency participants with reduced forms (Middle Left panel), Intermediate proficiency participants with unreduced forms (Middle Right panel), Advanced proficiency participants with reduced forms (Bottom Left panel), Advanced proficiency participants with unreduced forms (Bottom Right panel). Shades of blue indicate small pupil dilation; shades of green/yellow indicate large pupil dilation. The contour lines represent the pupil dilation values

predicted by the model.

Table 4.2: 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	2.869	5.412	0.530	0.59605
ProficiencyReduction: Basic.Unreduced	5.484	6.578	0.834	0.40442
ProficiencyReduction: Intermediate.Unreduced	14.219	6.083	2.337	< 0.05
ProficiencyReduction: Advanced.Unreduced	14.706	5.657	2.600	< 0.01
ProficiencyReduction: Basic.Reduced	16.278	6.769	2.405	< 0.05
ProficiencyReduction: Intermediate.Unreduced	13.991	6.303	2.220	< 0.05
Smooth terms	**edf**	**Ref.df**	**F-value**	**p-value**
Tensor product: Time and P-O Consistency Index: Advanced.Unreduced	12.126	13.966	4.522	< 0.0001
Tensor product: Time and P-O Consistency Index: Basic.Unreduced	9.962	11.163	4.922	< 0.0001
Tensor product: Time and P-O Consistency Index: Intermediate.Unreduced	13.610	15.451	6.955	< 0.0001
Tensor product: Time and P-O Consistency Index: Advanced.Reduced	12.776	14.751	4.532	< 0.0001
Tensor product: Time and P-O Consistency Index: Basic.Reduced	14.856	16.927	8.506	< 0.0001
Tensor product: Time and P-O Consistency Index: Intermediate.Reduced	8.391	8.680	14.813	< 0.0001
Smooth: Baseline Pupil Size	1.013	1.022	49.377	< 0.0001
Smooth: Trail Index	2.244	2.691	3.862	< 0.05
Smooth: Pupil Gaze Coordinates X and Y	10.873	13.879	3.8281	< 0.0001
Random effect: ParIDConsis over time	696.427	1240	1.426	< 0.0001
Random effect: ItemReduc over time	614.337	1627	1.535	< 0.0001

As indicated by the shades of colours and values on the contour lines, these plots reveal several important aspects of dilation: (1) Dilation peaks around 800 to 1200 ms for both reduced (Left panels in Figure 4.2) and unreduced forms (Right panels in Figure 4.2) and for all proficiency level participants (Top, Middle and Bottom panels in Figure 4.2) as seen in the grand averages plot (Figure 4.1). (2) Reduced forms appear to elicit an overall larger dilation and greater peak (Left panels in Figure 4.2) than unreduced counterparts (Right panels in Figure 4.2). (3) The basic proficiency participants showed a very different overall pattern/interaction between P-O consistency and time for reduced compared to unreduced forms. The basic proficiency participants with reduced forms exhibit the earliest and highest rise of dilation in the middle range, 0.6 to 0.7, of the index (Top Left panels in Figure 4.2). The rise begins around 300 ms (transitioning from green to yellow) and it reaches the peak around 800 ms (Top Left panel in Figure 4.2) but with unreduced forms, they demonstrate much smaller peak dilation (approximately half size of dilation) and it occurs in the upper range, 0.8 to 1, of the index (Top Right panel in Figure 4.2), and the rise reaches the peak around 900 ms (Top Right panel in Fig-

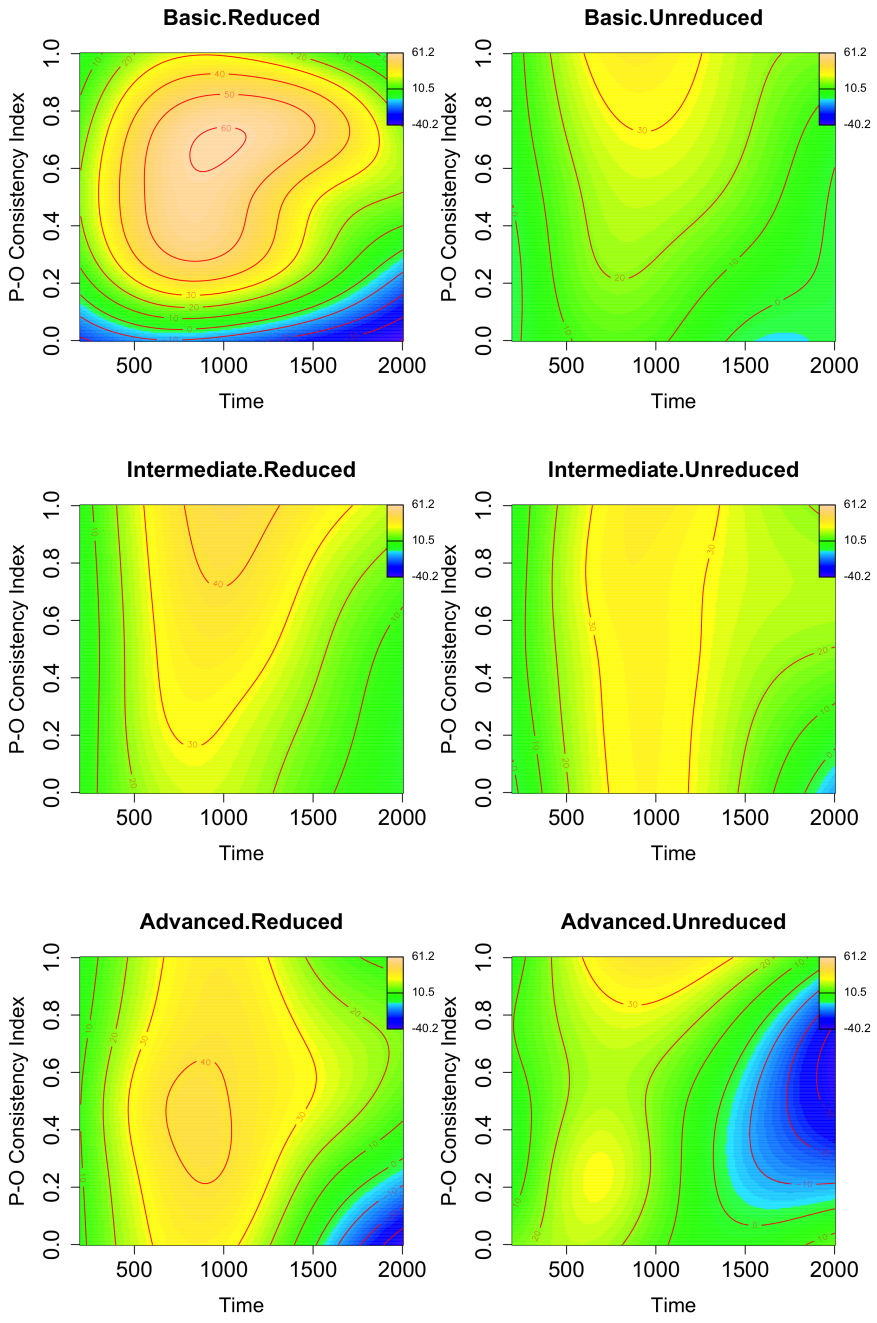

Figure 4.2: Contour plots of the interaction between the P-O Consistency Index, L2 Proficiency, and Reduction over Time: Basic proficiency participants with reduced forms (Top Left panel), Basic proficiency participants with unreduced forms (Top Right panel), Intermediate proficiency participants with reduced forms (Middle Left panel), Intermediate proficiency participants with unreduced forms (Middle Right panel), Advanced proficiency participants with reduced forms (Bottom Left panel), Advanced proficiency participants with unreduced forms (Bottom Right panel). Shades of blue indicate small pupil dilation; shades of green/yellow indicate large pupil dilation. The contour lines represent the pupil dilation values predicted by the model.

ure 4.2). (4) While the effect of P-O consistency is modest overall, reduced forms indicate a greater and more consistent P-O consistency effect around their peak dilation (Left panel in Figure 4.2), particularly with basic level participants (Top Left panel in Figure 4.2). Finally, (5) there is an overall trend of dilation becoming greater as the P-O consistency index increases (all panels in Figure 4.2), suggesting that consistent words induce greater dilation. This trend conflicts with the result of our study for L1 speakers and Hino et al. (2017) indicating that inconsistent words demonstrate slower response latency. Additionally, readily noticeable is that the difference between reduced and unreduced forms in the pattern of P-O consistency index are not uniform across L2 proficiency. The shapes of the effect appear to vary between the two forms, particularly along with the P-O consistency index for basic proficiency participants.

To further examine these aspects of the three-way interaction, we followed the procedure illustrated in Wieling (2018) and formally evaluated the difference between reduced and unreduced forms using the binary difference smooth. We first decomposed the tensor product into separate parts using a *ti* constructor: the effect of L2 proficiency over time, the effect of L2 proficiency over the P-O Consistency Index, and the interaction between Time, the P-O consistency index and L2 proficiency. We then re-specified the model with theses newly created binary variables representing if the word is reduced or not (see Wieling (2018) for an overview of a binary difference smooth and a *ti* constructor). The results demonstrate that the difference in pupil dilation between the reduced and unreduced forms over time reached significance only with basic proficiency participants ($edf = 2.752, F = 3.141, p < .05$) and the difference in pupil dilation between the two forms over the P-O consistency also reached significance only with basic proficiency participants ($edf = 3.081, F = 3.477, p < .05$), which verifies the points (3) and (4) above. The difference in pupil dilation between the two forms over the interaction between time and the P-O consistency reached significance with Advanced proficiency participants ($edf = 7.681, F = 2.077, p < .05$). With reduced forms, the modest effect of P-O consistency can be seen in the middle range of the index, 0.2 to 0.6, and the dilation rises slightly in the index range and falls around

the peak dilation, 800 to 1100 ms (Bottom Left panel), but with unreduced forms the small rise of dilation only occurs in the upper range of index, 0.8 to 1, and the steep falls occurs in the wide middle rage of the index, 0.2 to 0.8, after approximately 1400 ms (Bottom Right panel). That is, reduced forms for Advanced proficiency participants demonstrate a similar rise and fall patterns as other proficiency level participants with both forms to some extent, but unreduced forms for Advanced proficiency participants show a unique pattern, where the steep falls occurs in the wide middle rage of the index.

Overall, reduced forms appear to elicit an greater dilation than unreduced counterparts, and the basic proficiency participants showed a different overall pattern between P-O consistency and time for reduced compared to unreduced forms. While the effect of P-O consistency is modest overall, reduced forms indicate a greater and more consistent P-O consistency effect, particularly with basic level participants. There is an overall trend of dilation becoming greater as the P-O consistency index increases.

Delayed naming task

Similar to the Go-NoGo task, Figure 4.3 displays the grand average of pupillary responses over time for reduced and unreduced forms from -1500 ms to 2500 ms in the delayed naming task. Although the reduced form demonstrates a slightly greater dilation over time after 1000 ms, the time course of the dilation appears to be comparable between the two forms. That is, as time progresses, pupil dilation increases. The mean error rate for naming responses across participants was 6.43 % (SD = 5.25).

For the delayed naming task, we chose the time window from 200 ms to 2500 ms post stimulus onset for our analyses and employed the same variables and model fitting procedure as earlier. For random effects, we included 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, and for fixed effects, we eliminated Word Duration, Logged Target Word Frequency, Logged Number of Phonological Neighbours, and Standardized Num-

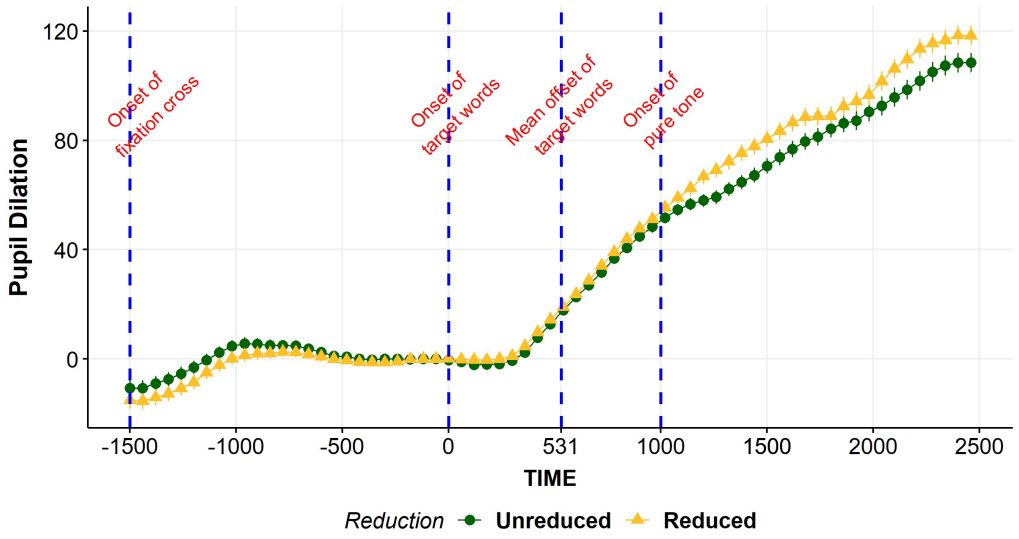

Figure 4.3: The grand average of pupillary responses over time for reduced and unreduced word forms in the delayed naming task. The vertical dot line at -1500 ms indicates the onset of the fixation cross, the line at 0 ms indicates the onset of stimuli, the line at 531 ms indicates the mean offset of stimuli, and the line at 1000 ms indicates the onset of pure tones.

ber of Homophones from the model during model fitting (the variables that remained in the model are illustrated in Table 4.3). We then added an AR-1 correlation parameter p = 0.987 and refitted the model with *scaled-t* family. Table 4.3 summarizes our final model. The parametric coefficients indicate that the overall pupil dilation is greater for Intermediate.Unreduced ($t = 2.337, p < 0.05$), Advanced.Reduced ($t = 2.600, p < 0.01$), Basic.Reduced ($t = 2.405, p < 0.05$), and Intermediate.Reduced ($t = 2.405, p < 0.05$) than Advanced.Unreduced (the reference level). The smooth terms reveal the significance of non-linear patterns associated with the input variables. We further discuss the summary of the final model together with visualization of the results. These contour plots in Figure 4.4 illustrates the interaction between the P-O Consistency Index, L2 Proficiency, and Reduction over Time: Basic proficiency participants with reduced forms (Top Left panel), Basic proficiency participants with unreduced forms (Top Right panel), Intermediate proficiency participants with reduced forms (Middle Left panel), Intermediate proficiency participants with unreduced forms (Middle Right panel), Advanced proficiency participants with reduced forms (Bottom Left panel), Advanced proficiency participants with unreduced forms (Bottom Right panel). Shades of blue

indicate small pupil dilation; shades of green/yellow indicate large pupil dilation. The contour lines represent the pupil dilation values predicted by the model.

Table 4.3: 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	61.054	6.707	9.104	< 0.0001
ProficiencyReduction: Basic.Unreduced	-23.783	7.675	-3.099	< 0.01
ProficiencyReduction: Intermediate.Unreduced	-22.159	7.275	-3.046	< 0.01
ProficiencyReduction: Advanced.Reduced	5.917	5.217	1.145	0.252
ProficiencyReduction: Basic.Reduced	-16.119	7.882	-2.045	< 0.05
ProficiencyReduction: Intermediate.Reduced	-14.286	7.387	-1.934	0.053
Smooth terms	**edf**	**Ref.df**	**F-value**	**p-value**
Tensor product: Time and P-O Consistency Index: Advanced.Unreduced	5.814	6.620	19.875	< 0.0001
Tensor product: Time and P-O Consistency Index: Basic.Unreduced	3.017	3.028	65.745	< 0.0001
Tensor product: Time and P-O Consistency Index: Intermediate.Unreduced	8.556	10.216	26.274	< 0.0001
Tensor product: Time and P-O Consistency Index: Advanced.Reduced	7.442	8.170	18.237	< 0.0001
Tensor product: Time and P-O Consistency Index: Basic.Reduced	9.494	11.964	18.990	< 0.0001
Tensor product: Time and P-O Consistency Index: Intermediate.Reduced	7.982	9.229	39.834	< 0.0001
Smooth: Baseline Pupil Size	3.388	4.212	27.432	< 0.0001
Smooth: Trail Index	4.321	5.261	28.545	< 0.0001
Smooth: Pupil Gaze Coordinates X and Y	17.746	18.795	49.844	< 0.0001
Random effect: ParIDConsis over time	710.868	1165	2.047	< 0.0001
Random effect: ItemReduc over time	535.711	2257	0.884	< 0.0001

These plots reveal the following aspects of the dilation: (1) Pupil dilation increases over time and reaches its peak at 2500 ms (all panels in Figure 4.4) as we saw in the grand averages plot (Figure 4.3). (2) Reduced forms appear to evoke a steeper, earlier, and higher rise of dilation (Left panels in Figure 4.4) than unreduced counterparts (Right panels in Figure 4.4), especially for Intermediate and Advanced proficiency participants (Middle and Bottom panels in Figure 4.4). (3) In contrast to the Go-NoGo task, the Advanced proficiency participants with reduced forms exhibit the steepest, earliest, and highest rise of dilation (Bottom Left panel in Figure 4.4). The rise begins around 800 ms (transitioning from blue to green) at the highest P-O consistency index and the rise starts around 1000ms at the lowest index at which the contour line represents the pupil dilation value of 80, and it reaches the peak dilation at 2500 ms (Bottom Left panel in Figure 4.4). (4) While the effect of P-O consistency is modest overall, reduced forms indicate a greater and more consistent P-O consistency effect around the peak dilation (Left panels in Figure 4.4), particularly with Advanced proficiency participants (Bottom

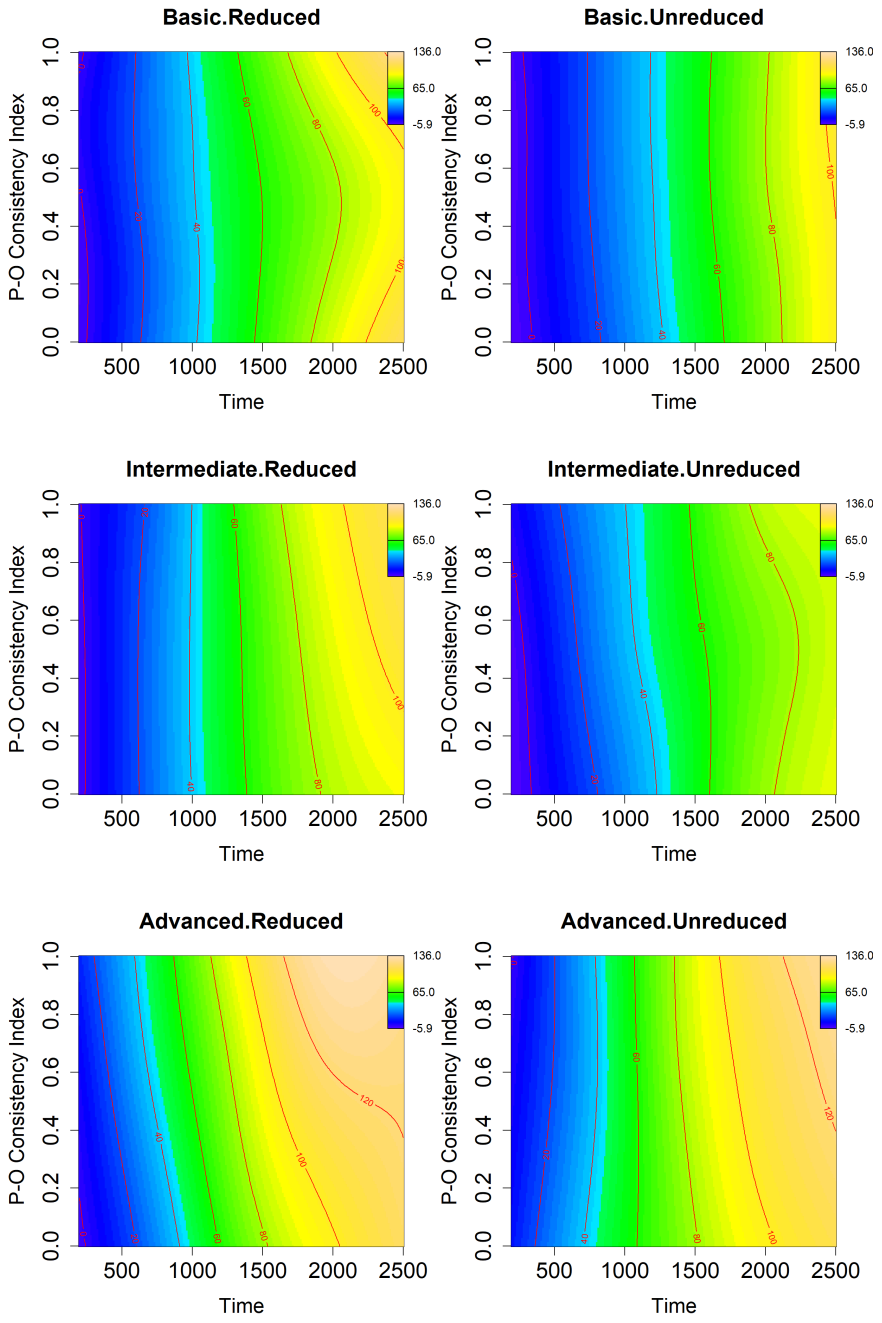

Figure 4.4: Contour plots of the interaction between the P-O Consistency Index, L2 Proficiency, and Reduction over Time: Basic Proficiency Level for Reduced Form (Top left panel), Basic Proficiency Level for Unreduced Form (Top Right panel), Intermediate Proficiency Level for Reduced Form (Middle Left panel), Intermediate Proficiency Level for Unreduced Form (Middle Right panel), Advanced Proficiency Level for Reduced Form (Bottom Left panel), Advanced Proficiency Level for Unreduced Form (Bottom Right panel). Shades of blue indicate small pupil dilation; shades of green/yellow indicate large pupil dilation. The contour lines represent the pupil dilation values predicted by the model.

Left panels in Figure 4.4), and finally, (5) similar to the Go-NoGo task, there is an overall trend where the dilation increases, as well as begins early, as the P-O consistency index increases (all panels in Figure 4.4). This trend is in line with the result of our study for L1 speakers for unreduced forms, but it differs from the previous literature demonstrating no effect of P-O consistency in auditory naming tasks (Rastle et al., 2011; Ziegler et al., 2004). Furthermore, the difference between reduced and unreduced forms in the pattern of dilation related to the P-O consistency index are comparable across L2 proficiency. The shapes of the effect appear to be similar across L2 proficiency levels, but the degree and timing of the P-O consistency effect differs between the two forms across L2 proficiency. That is, reduced forms exhibit an earlier and more clear effect of P-O consistency and it becomes more prominent as L2 proficiency increases. As in the results of the Go-NoGo task, we conducted further examination of the three-way interaction and formally evaluated the difference between reduced and unreduced forms using a binary difference smooth. The results demonstrate that the difference in dilation over time between the two forms reached significance with Intermediate and Advanced proficiency participants ($edf = 3.644, F = 4.957, p < .001$ for Intermediate and $edf = 2.179, F = 3.768, p < .05$ for Advanced proficiency participants) and the difference in dilation between the two forms over the P-O consistency reached significance with Advanced proficiency participants ($edf = 2.006, F = 3.715, p < .05$), which support the points (2), (3) and (4) above. The difference between the two forms in the interaction between time and P-O consistency reached significance with Intermediate proficiency participants ($edf = 2.448, F = 4.122, p < .05$). The modest effect of P-O consistency can be seen in the middle and upper range of the index, 0.3 to 1, with reduced forms around the peak dilation (Middle Left panel in Figure 4), but with unreduced forms the effect almost disappears. The dilation rises over time for both forms but the rise is smaller and more gradual for unreduced forms (Middle Right panel in Figure 4).

A question may arise with regard to whether our time window (200 to 2500 ms) is sufficient, since pupil dilation continues to rise and does not yet hit the peak at 2500 ms. In the task, the participants were expected to complete the naming

of the target word at around 2000 ms, as the offset of the pure tone is at 1500 ms and the expected onset of naming is at around 1500 ms and the expected offset of naming is at around 2000 ms (the mean duration of target words are 531 ms). This suggests that pupil dilation continues to rise after the completion of naming. However, Figure 4.3 seems to indicate that the dilation has peaked and will fall after 2500 ms.

In sum, reduced forms appear to evoke a greater rise of dilation than unreduced counterparts, especially for Advanced proficiency participants. While the effect of P-O consistency is modest overall, reduced forms indicate a greater and more consistent P-O consistency effect, particularly with Advanced proficiency participants, and similar to the Go-NoGO task, the dilation increases as the P-O consistency index increases.

4.4 General Discussion

In the present study, we investigated how the P-O consistency effect interacts with phonetic reduction for L2 Japanese speakers of L1 English. We used two types of tasks, Go-NoGo and delayed naming, in combination with the pupillometry technique, to compare the time course of the P-O consistency effect between reduced and unreduced Japanese words. We predicted an interaction between the effect of reduction and P-O consistency as we expected P-O consistency to affect reduced and unreduced forms differently. We also predicted an interaction between the effect of P-O consistency and L2 proficiency level, indicating that either the P-O consistency effect influences only high proficiency learners or only low proficiency learners. Our findings are as follows: (1) reduced forms elicited larger dilation than unreduced counterparts, (2) the P-O consistency effect emerged in both tasks, and it influences reduced and unreduced forms differently, (3) the consistency effect also varies depending on L2 proficiency, and finally (4) dilation becomes greater as the P-O consistency increases. In what follows, we discuss each aspect of the results of both tasks considering our predictions and previous literature.

First, as in previous studies (e.g., Arai, 1999; Ernestus et al., 2002; 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 a smaller pupil dilation. However, the difference between the two forms was not as large as expected (Figure 4.2 and 4.4). The difference between the two forms reached significance only with basic proficiency participants in the Go-NoGo task and with intermediate and advanced proficiency participants in the delayed naming task. This small difference between the two forms could be due to P-O consistency. There is an overall trend, as shown by the contour lines representing pupil dilation values in these plots, that the difference between reduced and unreduced forms is greater for consistent words (higher consistency values) than for inconsistent words (lower consistency values). That is, the processing cost of reduced forms is prominent in consistent words, but it attenuates or disappears in inconsistent words. For example, for the intermediate proficiency participants in the Go-NoGo task, the peak dilation in the higher range of P-O consistency is at the value of 40 in the reduced forms and 30 in the unreduced forms, but the peak dilation in the lower range of P-O consistency is at the value of 20 to 30 in reduced forms and 30 in unreduced forms (Middle Left and Right panels in Figure 4.2). For the advanced proficiency participants in the same task, the peak dilation in the higher range of P-O consistency is at the value of 30 in the reduced forms and 30 in the unreduced forms, but the peak dilation in the lower range of P-O consistency is at the value of 30 in the reduced forms and 20 in the unreduced forms (Bottom Left and Right panels in Figure 4.2). Similar patterns were also observed in the basic proficiency learners in the Go-NoGo task (Top Left and Right panels in Figure 4.2) and the intermediate and advanced learners in the delayed naming task (Middle and Bottom Left and Right panels in Figure 4.4). [3] These results suggest that P-O consistency plays an important role in the recognition difficulty of reduced word forms. This is in line with the discussion of orthographic effects in the recognition of reduced variants in previous studies (Bürki et al., 2012; Charoy & Samuel, 2019; Racine et al., 2014; Ranbom & Connine, 2007, 2011; Viebahn et al., 2018), arguing that the rea-

[3] As we predicted, overall pupillary responses were greater for the delayed naming task than for the Go-NoGo task, since a Go-NoGo task requires only passive listening for target items.

son that unreduced forms are easier to process than reduced counterparts could be due to the consistent relationship between the unreduced pronunciation and the orthographic form of words. However, our results appear to suggest that P-O consistency attenuates (or strengthens) the processing difficulty of reduced forms, rather than attenuating (or strengthening) the processing ease of unreduced forms. If unreduced forms are easier to process than reduced forms because of the consistent relationship between the unreduced pronunciation and orthography of words as suggested by previous studies, we should have observed the clear consistency effect in unreduced forms. However, we observed a clear consistency effect in reduced forms, indicating that P-O consistency influences the recognition difficulty of reduced forms. In contrast to the results for L1 speakers in Chapter 3, the interaction between P-O consistency and the processing cost of reduced forms was not in the expected direction, where consistent words were more difficult to process than inconsistent words. We need to further investigate this interpretation of our result by testing it against L1 speakers with alphabetic languages as the pronunciation and spelling mapping is more straightforward.

Recent research by Viebahn & Luce (2020) demonstrated that the processing difficulty of reduced forms that was found in a lexical decision task (slower response times) disappeared in a shadowing task, where participants listened and quickly repeated a stimulus. The researchers argue that meta-linguistic decision making processes in lexical decisions amplifies the effect of processing difficulty in reduced variants. Our result is in line with their results to some extent, but we found a small reduction effect in both delayed naming and Go-NoGo tasks. Moreover, although our Go-NoGo task does not require any linguistically motivated decision-making for the target words, we found the effect. This could be due to one of two factors: (1) this is because our participants were L2 speakers, but we found the reduction effect for L1 speakers as well in the third chapter, and (2) this is because we measured pupil dilation (online measure) rather than measuring accuracy rates and response latencies (offline measures). This result suggests that applications of online measures, such as pupillometry, should play an important role in the investigation of processing cost for reduced forms.

Second, we observed the effect of P-O consistency for both tasks. Previous studies demonstrated that while the effect was found in a Go-NoGo task, the effect was null (or reached significance only in by-subject analysis) in a naming task (Ventura et al., 2007, 2004; Ziegler et al., 2004). The important difference between our experiment and their experiment is the measurement utilized. Whereas these studies employed the offline measure, naming latency, we used online measure, pupil dilation. Our application of pupillometry (online measure) to a naming task has revealed that the consistency effect emerges in the task despite the fact that the effect was modest. Studies that found the consistency effect without meta-phonological or meta-linguistic tasks also employed online measures (Pattamadilok et al., 2009; Perre et al., 2011). Additionally, the timing at which the effect appears seems to be late in the delayed naming task, around 2000 ms in unreduced and 1500 ms in reduced forms. The results of this late effect is along the line of the discussion of the relative time course of the effect in Rastle et al. (2011). They argue that phonological representations are activated before the orthographic information comes in when the stimulus is a spoken word. When the task is simply to repeat the spoken stimulus as fast as possible, the phonological activation can drive the process of speech production before the orthographic information comes into effect. When the task requires additional processing stages, such as a post-perceptual decision stage for lexical decisions, the extra time provides an opportunity for inconsistent words to activate orthographic neighbours corresponding to conflicting phonological representations, thus slowing response times, or for consistent words to activate orthographic information that reinforces the correct phonological representation, thus speeding response times. Since our naming task was a delayed naming, there was an opportunity for the consistency effect to be exerted. The effect emerged within the time window of articulation (the offset of pure tone, at 2000 ms) rather than the time window of perception and planning (the onset of target words to the onset of pure tone, from 0 ms to 1000 ms). In addition, Rastle et al. (2011) also noted the possibility of task difficulty driving the activation of orthographic information, rather than the relative time course. Our results suggest that it is also an important factor for the effect because reduced forms were more

difficult to process than unreduced counterparts as evidenced by a greater pupil dilation, and reduced forms elicited an earlier and clearer effect of P-O consistency than unreduced counterparts. We speculate that information related to orthography and P-O consistency is not the cue that is utilized first during speech processing. Instead, we suggest that when additional information is needed to resolve difficulty of word recognition, these cues come into play. It seems that activation of orthographic information is automatic to some extent, as we observe the P-O consistency effect for both forms in both tasks. However, the activated information does not play an important role in the processing of unreduced forms as they are easy to process and listeners are not required to perform a deep level processing (e.g., phoneme level analysis) on them in both tasks. Importantly, our results indicate that the extent to which participants are required to process reduced (and unreduced) forms, as well as the extent to which the activated orthographic information is utilized (as indicated by the effect of P-O consistency) vary depending on the level of L2 proficiency.

Third, our results demonstrate the interaction between reduction (Reduced and Unreduced), L2 proficiency (Basic, Intermediate, and Advanced), and P-O consistency index (0: Inconsistent to 1: Consistent) over time. First, the difference in pupil dilation between the reduced and unreduced forms over time reached significance with Basic proficiency participants in the Go-NoGo task and with Intermediate and Advanced proficiency participants in the delayed naming task. The difference in pupil dilation between the two forms over the P-O consistency also reached significance with Basic proficiency participants in the Go-NoGo task and with Advanced proficiency participants in the delayed naming task. In the Go-NoGo task, the Basic proficiency participants exhibit a clear effect of consistency which weakens as the L2 proficiency level increases, and in the delayed naming task, we observe a pattern where the advanced proficiency participants exhibits a clear effect of consistency and it weakens as L2 proficiency level decreases. These patterns apply to both types of word forms but reduced forms show greater dilation and greater consistency effects overall. The results of the delayed naming task are in line with Veivo et al. (2016) showing that high proficiency level learners show

an effect of orthography but low proficiency level learners do not because of the high interaction efficacy between orthographic and phonological representations for high proficiency level learners. However, the consistency effect found in the Go-NoGo task displays a different direction, where the effect is strongest in basic proficiency participants and it weakens as L2 proficiency level increases. At first, this seems to contradict the results in Veivo et al. (2016); however, this could be due to the difference in tasks. In Veivo et al. (2016), orthographic information was visually presented to participants, meaning that the activation of orthography was given to participants without any cost, but Low and Intermediate proficiency participants' phonological and orthographic representations are less interactive due to their lack of development. As a result, the interference from the activated orthographic information was minimum. In our experiment, however, all information was provided auditorily, so that participants needed to access orthographic information via phonological representations. Since Basic (and Intermediate) proficiency level learners' phonological and orthographic representations are not yet well connected, it is effortful for the participants to access orthographic information via phonological representation. Additionally, since reduced forms are realized less efficiently, orthographic information was needed to be exerted to increase efficacy in the realization, even though it was cognitively demanding. This doubly effortful process leads to greater dilation and strong P-O consistency effects, and as their proficiency increases, the effort to access orthographic information via phonological representation becomes easier; therefore, the dilation becomes less and the effect of P-O consistency becomes small.

We could extend our analysis on the reaction times for GO trials (i.e., a button press for a pure tone). The reaction time could be used to evaluate the degree of participants' attentiveness to the task. The fast (more attentive) and slow reaction time (less attentive) participants might show different patterns of interactions of these effects because less attentive participants may have not made any effort, which would be indicated by overall weak pupillary responses, thereby diminishing these effects. Furthermore, L2 learners tend to show great variability in vocabulary size. It is therefore likely that our participants encountered words that they

were not familiar with while performing the task. This does not prevent them from successfully performing the task, but it may affect the degree of activation of orthographic information. Previous research has shown that there is no orthographic consistency effect for pseudowords (Pattamadilok, Morais, Colin & Kolinsky, 2014; Ventura et al., 2004). This result could apply to unfamiliar L2 words as well. Further research is needed to investigate the activation of orthography for unfamiliar words for L2 speakers.

Fourth, the direction of the consistency effect found in our results was opposite to what has been found in Hino et al. (2017) and other previous studies (e.g., Ziegler & Ferrand, 1998; Ziegler et al., 2004). Similar patterns were found in the results of experiments in L1 speakers for unreduced forms in the third chapter. As discussed in the third chapter, this could be caused by correlations between the consistency effect and other lexical variables. For example, P-O inconsistent words tend to have few orthographic neighbors and low-frequency bigrams (Ziegler, Petrova & Ferrand, 2008). Further analysis of our results have revealed that low P-O consistency words tend to have a higher number of phonological neighbours. Yoneyama (2002) shows that high phonological neighbourhood density facilitates the recognition of Japanese words, meaning that the facilitatory effect of high phonological neighbourhood density competes against the inhibitory effect of low P-O consistency. For further research, we could use pseudowords to control for this confounding effect. Another possible interpretation of the opposite direction of the consistency effect is that the direction of the P-O consistency effect changes depending on the type of orthography (logographic or alphabetic). As discussed in the third chapter, previous research has shown that in French (alphabetic), inconsistent words elicited increased negativity (N400) (Pattamadilok et al., 2009; Perre et al., 2011), but in Chinese (logographic), the effect is reversed such that consistent words elicited increased negativity (Chen et al., 2016). That is, the direction of P-O consistency effect differs between logographic and alphabetic languages. While much research has been done cross-linguistically, less research, if any, investigates this directionality issue. We need further research to compare the direction of the effect cross-linguistically.

Importantly, our application of pupillometry to the study of both P-O consistency and reduced speech is novel. While the measurement of pupil dilation has been used to investigate various aspects of spoken language, our study represents the applicability of the method to examine such effects in spoken word recognition. In addition to the online measures that have been employed in the previous literature (e.g., eye-tracking and EEG), pupillometry allows us to investigate the degree and time course of such effects without an overt task or behavioral response. This is particularly beneficial for participants who would have difficulty performing a specific task (Laeng et al., 2012), such as pre-verbal children and people with aphasia.

Lastly, research on the P-O consistency effect has important implications for L2 acquisition. As discussed in the introduction section, there might be unbalanced lexical knowledge in terms of phonological and orthographic information in L2 speakers due to the early introduction of orthography in L2 learning, a strong orthographic component in the learners' lexical knowledge (Veivo & Järvikivi, 2013; Veivo et al., 2016). Previous research has shown that orthography can be advantageous or disadvantageous in L2 acquisition depending on the aspect of the language and the learner's proficiency levels (e.g., Escudero et al., 2008; Escudero, Simon & Mulak, 2014; Hayes-Harb et al., 2018; Hayes-Harb, Nicol & Barker, 2010). As suggested by Hayes-Harb et al. (2018), it is important to further investigate the factors that influence the beneficial or detrimental effects of written input and the role that instruction may play in learners' use of written L2 input. This could be particularly crucial for learners whose L1 writing system is incongruent with L2 writing systems (e.g., alphabetic vs. logographic). Moreover, we need to further investigate how these strategic choices made by learners and/or instructors (e.g., use or non-use of early written input) predict the ultimate attainment of their L2 acquisition.

4.5 Conclusion

Our study has demonstrated that the effect of P-O consistency influences reduced and unreduced word forms in different ways. While the consistency effect emerged for both forms, the effect was more consistent and clearer for reduced forms than for unreduced forms. Also, the processing cost of reduced forms varied depending on P-O consistency, where consistent words were more difficult to process than inconsistent words. The consistency effect is also dependent on both L2 proficiency level and task type, where high proficiency learners show the strong effect of consistency and the effect becomes weaker as L2 proficiency decreases in the delayed naming task, but low proficiency learners show the strong effect of consistency and the effect becomes weaker as L2 proficiency increases in the Go-NoGo task. We have also revealed that P-O consistency plays an important role in the recognition difficulty of reduced forms. Finally, we showed that pupillometry can be used as an informative research tool to investigate the effect of P-O consistency and reduced speech.

Chapter 5

General discussion and conclusion

The objective of this dissertation was to investigate phonetic variability of word-medial voiced stops and word-medial nasals in Japanese across different styles of speech and how such variability interacts with phonological-orthographic (P-O) consistency in spoken word recognition in L1 and L2 Japanese. With one corpus analysis and two pupillometry experiments, we found a gradual effect of reduction across speech styles, where the most spontaneous speech demonstrates the greatest reduction, as well as a comparable reduction pattern between Japanese and English. We also found dynamic interactions among reduction, P-O consistency and L2 proficiency, where first, reduced forms elicited larger dilation than unreduced counterparts, second, the P-O consistency effect influenced reduced and unreduced forms differently, and third, the consistency effect also varied depending on L2 proficiency. Additionally, pupillary studies demonstrated the applicability of pupillometry to examine such effects in spoken word recognition. In what follows, we first briefly summarize the results of the individual studies and discuss how the findings of the three studies interact with each other and what the implications of these results are. This is followed by a discussion of limitations and suggestions for future studies.

5.1 Summary of results

Study 1 (Chapter 2) investigated the distribution and degree of phonetic reduction across various styles of speech using the Corpus of Spontaneous Japanese

(Maekawa, 2003). Specifically, we examined the effect of speech style on the realization of word-medial stops and word-medial nasals and discussed the acoustic results of phonetic reduction in comparison to Warner & Tucker (2011), who investigated the effect of speech style on the realization of word-medial stops (and flaps) in English. The present results revealed a gradual effect of reduction across speech styles specifically when using intensity difference as a measure of reduction. The effect indicated a tendency of predicted reduction patterns, i.e., the most spontaneous speech demonstrates the greatest reduction, and the least spontaneous speech exhibits the smallest reduction. However, we did not find a clear gradual effect as suggested by Maekawa (2005). That is, although we found the tendency towards the predicted reduction patterns across the speech styles, the overall effect was more categorical; Dialogue (most spontaneous) shows greater reduction than Read Speech (the least spontaneous). When we employed an intensity difference as a measure of reduction, the acoustic results of reduction and their distributional patterns across speech styles for voiced stops were most comparable between Japanese in the present study and English in Warner & Tucker (2011), and nasals also indicated comparable reduction patterns. These results extend the findings of Barry & Andreeva (2001) that there are language-independent patterns of reduction.

Using pupillometry, Study 2 (Chapter 3) examined how P-O consistency interacts with reduced word forms in Japanese by comparing the time course of the consistency effect between reduced and unreduced word forms in Go-NoGo and delayed naming tasks with native Japanese listeners. We found that reduced word forms elicited greater pupil dilation than unreduced word forms in both tasks, indicating a greater processing load for reduced forms than for unreduced forms. For the Go-NoGo task, we found an effect of P-O consistency for both forms, but the direction of the effects differed between the two forms, suggesting that the consistency effect influenced the two forms differently. Our results also indicated that regardless of the explicitness of the task, orthography (P-O consistency) plays a role for the comprehension of careful speech (unreduced forms). In contrast to the results of Mitterer & Reinisch (2015), we found the effect of orthography in

an implicit task with casual speech (reduced forms). This 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 their study, these forms were presented in isolation in our study. This result suggests that while the orthography plays a role in the recognition of reduced forms in isolation, the effect disappears when such forms are presented with more conversation-like context, suggesting that the way in which reduced forms are presented plays a more important role than the explicitness of tasks. For the delayed naming task, while we observed the P-O consistency effect for both forms, the effect was greater for reduced forms. This suggests that the P-O consistency effect influenced the two forms differently and that the consistency effect played an important role in the processing cost of reduced forms. The direction of the consistency effect found in unreduced forms for the Go-NoGo task was opposite to what has been found in Hino et al. (2017) and other previous studies (e.g., Ziegler & Ferrand, 1998; Ziegler et al., 2004). This could be due to correlations between a low P-O consistency and higher phonological neighbourhood density, or by orthography-specific P-O consistency effects: alphabetic (Pattamadilok et al., 2009; Perre et al., 2011) and logographic (Chen et al., 2016).

Study 3 (Chapter 4) extended the investigation of Study 2, examining how the P-O consistency effect interacts with phonetic reduction for learners of Japanese with English as their L1. As in Study 2, we used two types of tasks, Go-NoGo and delayed naming, paired with pupillometry, to compare the time course of the P-O consistency effect between reduced and unreduced Japanese words. We observed that unreduced forms are easier to process than their reduced counterparts in both tasks. However, the difference between the two forms was not as large as expected. This could be due to P-O consistency as we found an overall trend that the difference between the two forms is greater for consistent words than for inconsistent words. These results suggest that P-O consistency plays an important role in the processing cost of reduced word forms, which is in line with the discussion of orthographic effects in the recognition of reduced variants (e.g., Racine et al., 2014; Ranbom & Connine, 2007, 2011). However, in contrast to previous studies and

Study 2 (Chapter 4), consistent words were more difficult to process than inconsistent words. In addition, we observed the effect of P-O consistency for both tasks although previous studies demonstrated that the effect was null in a naming task (Ventura et al., 2007, 2004; Ziegler et al., 2004). This difference could be because we used pupillometry, an online measurement. Furthermore, our results demonstrated an interaction between Reduction, L2 proficiency, and P-O consistency over Time. However, the interaction pattern found in Go-NoGo task displays a different direction compared to what has been found in Veivo et al. (2016). This could be due to the difference in tasks, where orthographic information was visually presented to participants in Veivo et al. (2016) but it was not in our study. Finally, similar to the results of Study 2, the direction of the consistency effect found in our results was opposite to what has been found in previous studies (e.g., Hino et al., 2017; Ziegler et al., 2004). A possible reason for this could be, as discussed for Study 2, correlations between a low P-O consistency and higher phonological neighbourhood density, or orthography-specific P-O consistency effects differing between alphabetic (Pattamadilok et al., 2009; Perre et al., 2011) and logographic (Chen et al., 2016) systems.

5.2 General discussion

Throughout this dissertation, we investigated how speech production varies across different styles of speech, how listeners interact with such produced variability, and how orthographic information interacts with this variability during spoken language comprehension. We found a gradual effect of reduction across speech styles and a comparable reduction pattern between Japanese and English. We also found dynamic interactions among reduction, P-O consistency and L2 proficiency, as well as the applicability of pupillometry to examine such effects in spoken word recognition. In this section, we discuss how the results of these studies interact and what the implications of these results are. Our discussion will be organized by the following sections: (1) Phonetic variability in speech production, (2) Importance and applications of research on production of reduced speech, (3) P-O consistency

and the role of reduced speech in perception, (4) Importance and applications of orthographic effects in spoken word recognition models, (5) Methodological considerations of research on P-O consistency and the role of reduced speech in perception.

5.2.1 Phonetic variability in speech production

In Study 1, we investigated how speech production varies across different styles of speech. Over the years, research on phonetic variability and spontaneous speech has increased and revealed many pronunciation variants, including changes in acoustic properties of segments and disappearances of expected segments (e.g., Dilts, 2013; Ernestus, 2000; Johnson, 2004). Such phenomena in the realization of words in spontaneous casual speech are frequent and usually accounted for by a gradual process of phonetic reduction. While some studies, including Study 1, treat reduction and deletion of segments separately to investigate a particular reduction phenomena (Turnbull, 2018), most researchers view phonetic variability as phonetic surface forms presenting a continuum of realizations ranging from forms with temporally and spectrally reduced to forms with total disappearance (or with subtle remnant) of acoustic properties (e.g., Davidson, 2011; Manuel, Shattuck-Hufnagel, Huffman, Stevens, Carlson & Hunnicutt, 1992). Furthermore, gradient reduction processes and their interactions with a style of speech have been documented in various languages. Study 1 extended the results of Arai (1999) and Warner & Tucker (2011) by demonstrating a gradual effect of reduction in word-medial voiced stops and word medial-nasals across various types of speech in Japanese and a comparable reduction pattern of voiced stops between English and Japanese. This result also provides evidence for the hypothesis in Barry & Andreeva (2001) that there are cross-language similarities in spontaneous speech patterns by demonstrating a comparable reduction pattern of voiced stops between English and Japanese across speech styles, particularly when measured by intensity difference.

5.2.2 Importance and applications of research on reduced speech

As a number of studies have pointed out (Ernestus & Warner, 2011; Tucker & Ernestus, 2016; Warner, 2011, 2012), reduced speech is of importance to understand how speech communication works. In particular, reduced speech tells us about the most common form of language that speech communication system interacts with. Importantly, a recent rapid advancement of speech technology has allowed us to approach reduced speech in a gradient and quantitative manner and apply in-depth and large-scale analysis motivated by frameworks/hypotheses that account for reduction phenomena, such as Hyper/Hypo theory (Lindblom, 1990), Smooth Signal Redundancy Hypothesis (Aylett & Turk, 2004), and Articulatory Phonology (Browman & Goldstein, 1990, 1992). Additionally, as a result of the development in computer technology, many large-scale speech databases have been created and become available for use, many of which are phonetically/phonemically labeled using forced-alignment. Taking Japanese speech corpora as an example, Maekawa (2015) provides an overview of various types of Japanese speech corpora, distinguishing (1) speaker types (adult vs. infant), (2) spontaneity (spontaneous vs. read), and (3) mode of speech (monologue vs. dialogue). These speech corpora can be used as representatives of various types of speech (Cangemi & Niebuhr, 2018). In particular, the CSJ dataset that we used for Study 1 can be a representative of both spontaneous and read speech, allowing us to compare differences and variability in phonetic surface forms between spontaneous casual speech that we deal with in natural interactions and careful laboratory speech, a strictly controlled and idealized form of language. Importantly, Akita & Kawahara (2005) built a generalized statistical model that could predict the occurrence probabilities of various reduced word-forms from the input of the base/unreduced word-form. They used the CSJ to build the model and showed that the resulting pronunciation dictionary containing reduced word forms improved the performance of automatic speech recognition systems for spontaneous speech.

The importance in the understanding of the most common form of language applies not only to research but also to applications that are closely related to our

life, including automatic speech recognition (Speech to Text (STT)), speech synthesis (Text to Speech (TTS)), and second language teaching. For example, recently the technology of online STT is included in activities we perform on a daily basis, such as watching a video on YouTube, having an online meeting on Zoom or Google Meet, giving a presentation with Microsoft PowerPoint or Google Slides, and creating a document with Google Document or Microsoft Word. The importance of coping with reduced speech has been increasing, with a focus on spontaneous conversational speech in language modeling (Zellers, Schuppler & Clayards, 2018) because acoustic models estimated with read speech data performed poorly on spontaneous speech (Adda-Decker & Lamel, 2018). Another example of application for reduced speech is second language teaching. Several studies suggest that it might be helpful to include reduced speech in second language teaching so that students can recognize the range of pronunciation variants in both production and perception. Only acquiring learner-directed exceptionally careful speech might cause students difficulty understanding natural speech outside the classroom (Shockey, 2003; Warner, 2011; Warner & Tucker, 2011). In short, Study 1 highlights phonetic variability of speech production across various types of speech and the importance of incorporating reduced forms in such applications.

5.2.3 P-O consistency and the role of reduced speech in perception

In this dissertation, not only did we investigate how production of speech varies but also how listeners interact with such variability. Research shows that these reduced variants are more difficult to process than unreduced counterparts despite the fact that such variants occur more frequently than canonical/unreduced ones (e.g., Arai et al., 2007; Ernestus et al., 2002; Tucker, 2007, 2011; van de Ven et al., 2011). Some researchers have argued that the reason that unreduced forms are easier to process than reduced counterparts could be because of the orthography, particularly the consistent relationship between the unreduced pronunciation and its orthographic form (Charoy & Samuel, 2019; Racine et al., 2014; Ranbom & Connine, 2007, 2011; Viebahn et al., 2018), suggesting that phonological and or-

thographic consistency could play an important role in the processing of reduced forms. In Study 2 and 3, we investigated how the effect of P-O consistency interacts with reduced word forms in L1 and L2 Japanese by comparing the time course of the consistency effect between reduced and unreduced word forms in a Go-NoGo and delayed naming task. We found first that reduced forms appeared to elicited larger dilation than their unreduced counterparts in both tasks, second that the P-O consistency effect emerged in both tasks and it influences reduced and unreduced forms differently, and third that dilation becomes greater as the P-O consistency decreases for L1 speakers (Study 2), but for L2 speakers, dilation becomes greater as the P-O consistency increases (Study 3).

Mitterer & Reinisch (2015) did not find an interaction between the processing cost of reduction and the type of deleted segment (whether orthographically represented or not) in their eye-tracking experiment. This result is in line with the discussion in Viebahn et al. (2018) suggesting that both phonological variants and spelling-sound consistency affect the processing of spoken novel words, but the effect of phonological variation overshadows the effect of spelling-sound consistency. Importantly, what we found in Study 3 extends their discussions by having the P-O consistency effect on a scale (0 to 1) rather than binary (consistent/presence or inconsistent/absence of a target segment). In Study 3 (L2 listeners), we found an interaction between the processing cost of reduction and P-O consistency. We observed that unreduced forms are easier to process than their reduced counterparts in both tasks. However, the difference between reduced and unreduced forms is greater for consistent words (higher consistency values) than for inconsistent words (lower consistency values). That is, the processing cost of reduced forms is prominent in consistent words, but it attenuates or disappears in inconsistent words. While these results provide evidence that P-O consistency plays an important role in the processing cost of reduced word forms, the direction of the interaction differed from what is discussed in the previous studies (Racine et al., 2014; Ranbom & Connine, 2007, 2011). While this interaction effect could be due to a strong orthographic component in L2 learners' lexical knowledge (Veivo et al., 2016), a similar pattern of the effect has also appeared in Study 2. While we found

a clear effect of reduction in both tasks, the effect of P-O consistency was modest in both tasks, particularly for the Go-NoGo task, which could be due to the fact that an orthographic component in L1 listeners' lexical knowledge is not as strong as L2 listeners' lexical knowledge, since phonological representations are established before orthographic information comes into play (Viebahn et al., 2018). The delayed naming task exhibited a clear effect of consistency, particularly in reduced forms, where steep pupil dilation for low P-O consistency words than for high P-O consistency words. The difference between reduced and unreduced forms was greater for inconsistent words (low consistency values) than for consistent words (high consistency values). That is, the processing cost in reduced forms is prominent in inconsistent words, but it attenuates or disappears in consistent words. This result suggests that P-O consistency plays an important role in the processing cost of reduced word forms by attenuating the cost with consistent orthographies, which is in line with the discussion of the previous studies (Racine et al., 2014; Ranbom & Connine, 2007, 2011).

Interestingly, the direction of the interaction between reduction and P-O consistency varies between L1 and L2 listeners. Whereas the processing cost of reduced forms is attenuated in inconsistent words in L2, the cost is attenuated in consistent words in L1. While both results indicate the importance of P-O consistency in the processing cost of reduced word forms, the result in L1 is more in line with the discussion of previous research, where a consistent relationship between pronunciation and orthography facilitates spoken word comprehension (Racine et al., 2014; Ranbom & Connine, 2007, 2011). This directional difference of the interaction between L1 and L2 listeners could be due to the L1 orthography of L2 speakers, as a number of studies show that L1 orthography affects L2 orthographic effects (Dornbusch, 2012; Veivo & Järvikivi, 2013; Veivo et al., 2016). Our L2 listeners are native English speakers, meaning that orthographic system between L1 (English) and L2 (Japanese) differs substantially (alphabetic vs. logographic). According to the psycholingusitc grain size theory, there is a different degree of phonological awareness for a different grain size of unit, and for English, a rhyme is a salient unit of mapping process between phonology and orthography (Ziegler & Goswami,

2005). For L2 (Japanese) however, a mapping process between phonology and orthography is abstract and we calculated P-O consistency based on the number of phonological and orthographic neighbours (Hino et al., 2017). This substantially difference way to map between phonology and orthography, as well as measuring the degree of phonological and orthographic consistency, might mean that even for advanced learners, they may not be able to access to or use the information of P-O consistency during spoken word comprehension. The consistency effect found for L2 listeners might have been the effect of phonological neighbourhood density, as low P-O consistency words tend to have a higher number of phonological neighbours, and high phonological neighbourhood density facilitates the recognition of Japanese words (Yoneyama, 2002). In order to investigate this issue, we need further research with heritage speakers of Japanese, whose Japanese proficiency is as high as native speakers.

5.2.4 Importance and applications of orthographic effects in spoken word recognition models

A number of studies on P-O consistency effects in spoken word recognition, as well as research on perception of reduced speech, discuss the importance of implementation of orthographic effects in a spoken word recognition model (e.g., Dornbusch, 2012; Ranbom & Connine, 2007, 2011; Ziegler et al., 2004, 2003). Our studies (Chapter 3 and 4) also highlighted the importance of orthographic (and P-O consistency) effects in spoken word recognition by demonstrating the interplay between phonology and orthography, particularly for reduced word forms. A general framework for word processing presented by McClelland & Rumelhart (1981) and Rumelhart & McClelland (1982) describes links between auditory and visual word processing. The framework predicts the influence of both phonological knowledge in visual word recognition and orthographic knowledge in spoken word recognition, demonstrating a close linkage between phonological and orthographic representations. While dual route models of visual word recognition only incorporate phonological activation from orthography (Coltheart, Curtis, Atkins & Haller, 1993; Coltheart, Rastle, Perry, Langdon & Ziegler, 2001), the bi-

modal interactive activation model (BIAM) can formalize orthographic activation from phonology (Grainger & Ferrand, 1996; Grainger, Spinelli, Farioli, Diependaele & Ferrand, 2003), where orthographic and phonological representations interact at both lexical and sublexical levels to influence spoken word recognition. Furthermore, Ranbom & Connine (2011) describe how the model operates with a mismatch between phonological and orthographic forms and how it is compatible with the restructuring view (Ziegler et al., 2004, 2003) accounting for the orthographic effects in such a way that orthographic knowledge modifies existing phonological representations by creating more specified phonological representations with orthographic information. Moreover, several studies argue that orthographic knowledge not only modifies existing phonological representations, but can create an additional phonological representation that contains features or segments that are not present in the spoken form (Connine, 2004; Connine, Ranbom & Patterson, 2008; Ranbom & Connine, 2007, 2011; Taft, 2006). We need further research to discuss a role of orthography during spoken word recognition with reference to the BIAM. Furthermore, while a few studies have developed computational models of dual route/process models (Perry, Ziegler & Zorzi, 2007, 2010) and the BIAM (Diependaele, Ziegler & Grainger, 2010) integrated with phonological representations for visual word recognition, there seems to be no computational models for spoken word recognition that include a role of orthographic representations (Hannagan, Magnuson & Grainger, 2013). Therefore, we also need the implementation of the BIAM to develop a computational model for spoken word recognition that includes the influence of orthography.

5.2.5 Methodological considerations of research on P-O consistency and the role of reduced speech in perception

This dissertation also raised a number of methodological issues. One of the issues discussed in Study 2 and 3 was the relationship between a type of speech and task. As pointed out by Mitterer & Reinisch (2015), as well as other studies (e.g., Cutler et al., 2010; Viebahn & Luce, 2020; Ziegler & Ferrand, 1998), the effect of orthography (e.g., P-O consistency) and speech style (reduced or unreduced word

forms) appear to interact with the type of task that participants perform. Mitterer & Reinisch (2015) noted that the contrast between the two experiments in their study confounded speech type with task. That is, they found the orthography effect in an explicit task (pronunciation judgement task) with careful speech (unreduced forms in isolated context) but they did not find the effect in an implicit task (visual world paradigm) with casual speech (reduced forms in conversational-like 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. In Study 2, we found the P-O consistency effect in both implicit tasks (Go-NoGo and delayed naming task) with reduced and unreduced forms, revealing two important points: (1) P-O consistency (and orthography) plays a role in the comprehension of careful speech (unreduced forms in isolation) in implicit tasks. This result extends the findings in Mitterer & Reinisch (2015) by revealing that there is an orthographic effect in an implicit task with careful speech, and (2) in contrast to the results of Mitterer & Reinisch (2015), we found the effect of P-O consistency (and orthography) in implicit tasks with casual speech (reduced forms in isolation). As discussed in Study 2, this could be because of the difference in the degree of "conversational-likeness". While reduced forms are presented with informal sentences including discourse markers and contractions in Mitterer & Reinisch (2015), items are presented in isolation in our study. This result suggests that while the orthography plays a role in the recognition of reduced forms in isolation, the effect disappears when such forms are presented with more conversation-like context, suggesting that the way in which words are presented likely plays a more important role than the types (explicitness) of tasks. The importance of ways in which such information is presented was also discussed in Study 3.

In Study 3, we found an interaction effect between Reduction, L2 proficiency, and P-O consistency over Time. While the interaction patterns found in the delayed naming task were in line with Veivo et al. (2016), the pattern found in the Go-NoGo task exhibited a different direction. In the delayed naming task, high proficiency learners showed an effect of consistency that weakens as L2 proficiency decreases. However, in the Go-NoGo task, the consistency effect was strongest for

basic proficiency participants and it weakened as L2 proficiency level increased. We discussed this difference with regard to a methodological difference in these tasks. That is, the ways in which orthographic information was presented to participants differed between the Go-NoGo task in our study and the cross-modal priming in Veivo et al. (2016), and the difference led participants to taking a different path to access orthographic information, resulting in a different direction of the interaction effect. In short, how a certain types of linguistic information is presented to participants in a given task plays an important role in how an effect of interest emerges.

This line of discussion can be applied to the comparison between our results in Study 2 and the results in Mitterer & Reinisch (2015) in order to extend the understanding of the relationship between the effect of P-O consistency and reduced speech. We found the P-O consistency (and orthography) effect in both implicit tasks (Go-NoGo and delayed naming task) with careful (reduced) and casual speech (unreduced forms), suggesting that P-O consistency (and orthography) plays a role in comprehension of both careful and casual speech. However, Mitterer & Reinisch (2015) did not find an effect of orthography in an implicit task (visual world paradigm) with casual speech. They only found the effect in an explicit task (pronunciation judgement task) with careful speech. They then concluded that orthography does not play a role in comprehension of casual speech. However, we postulate an alternative way to interpret their results. The reason that they found the effect in the pronunciation judgement task could be due to how the task was executed. In the pronunciation task, target words were visually presented in addition to auditory stimuli, and participants were not instructed to respond as fast as possible for their judgement of pronunciation. That is, visually presented orthography could activate orthographic information of target segments, and the meta-linguistic analysis of phonological and orthographic forms of target words prompted by pronunciation judgement tasks without time constraint could amplify the degree of activation of orthography. As a result, it is possible that the task could have imposed the effect of orthography. In short, the result of orthographic effects in Mitterer & Reinisch (2015) might have changed (there might have been a

null effect of orthography) if the task was executed differently.

Furthermore, we speculate an additional reason that Mitterer & Reinisch (2015) could have obtained a null effect for both tasks, which is the use of German and Maltese. P-O consistency plays an important role in the effect of orthography (Ziegler et al., 2003) for the judgement of pronunciation (Ranbom & Connine, 2011). While there seems to be no information about orthographic depth of Maltese, German is one of the languages that has a shallow orthography despite that fact that P-O consistency tends to be less consistent than O-P consistency in these languages (Landerl, 2005; Neef & Balestra, 2011; Seymour, Aro & Erskine, 2003). This suggests that German speakers might not be as sensitive to the effect of P-O consistency as speakers of English or French whose languages have deep orthographies. Moreover, according to the psycholingusitc grain size theory, there is a different degree of phonological awareness for a different grain size of unit, and a rhyme emerges as a salient grain size unit of processing in languages with complex phonological structures, such as English and German (Ziegler & Goswami, 2005), suggesting that the phoneme-grapheme mismatch created by deletion of target segments in German in Mitterer & Reinisch (2015) might not be as salient as the syllable-grapheme mismatch that has been studied in English and French in previous studies (e.g., Ziegler et al., 1997).

We need more studies on P-O consistency, particularly for Japanese, since it is not straightforward to measure P-O consistency in logographic languages, especially for Japanese, due to the use of multiple types of scripts (one logographic *Kanji* and two syllabic scripts *Hiragana* and *Katakana*). Future research on this issue will further our understanding of the nature and dynamics of orthographic information in a lexical representation and how it interacts with phonological representations during speech comprehension.

5.3 Limitations and future research

In what follows, we will point out several limitations of our studies and discuss how to address them in future research.

First, we made an assumption that orthographic representations are associated with logographic scripts in Japanese. Despite the fact that these words are normally written in logographic scripts, they can also be written in syllabic scripts because Japanese allows for the use of three types of scripts (logogram: Kanji and syllabaries: Hiragana and Katakana). Pylkkänen & Okano (2010) argue that orthographic representations are sound-based, script-invariant. They conducted a masked priming experiment with lexical decisions using two syllabaries in Japanese, Hiragana and Katakana. In the repetition prime condition, participants saw prime and target pairs in the same script, whereas in the script-change prime condition, they saw the pairs in different scripts. Their results indicated that regardless of script types (typical or atypical), there was a priming effect for both conditions. To assure that there was no semantic effect in this experiment, they carried out an unmasked priming experiment using the same prime and target pairs. The results showed a semantic priming effect which the masked priming experiment did not show, verifying that there was no semantic effect in the masked experiment. Pylkkänen & Okano (2010) also employed a magnetoencephalography (MEG) with lexical decisions. They divided their stimuli into high and low frequency words and presented them to participants in both typical and atypical scripts. The results revealed that the timing at which frequency effects emerged was the same regardless of script types. Additionally, Maurer, Zevin & Mccandliss (2008) found that the processing of visually presented typically and atypically spelled Japanese words elicits a similar N170 component of the ERPs.

These results raise a question regarding the nature of orthographic representation in Japanese, namely whether orthographic representations contain information of all possible forms of orthography, or whether they are only associated with a typical form of orthography. On the basis of the visual word comprehension studies reviewed above, the former seems to be the case. This issue is important for the investigation of P-O consistency effects because when we calculate P-O consistency, we use a typical forms of orthography, but if we also consider an atypical form of orthography, P-O consistency differs between a single word written in a typical form and atypical form of orthography. A possible experiment design to

investigate this issue is as follows. We conduct an eye-tracking experiment to identify what kind of orthographic information is activated and at what point in time the activation occurs during the perception of spoken words, as participants match spoken targets with printed words on a computer screen. If orthographic representations are only associated with a typical form of orthography, the timing at which lexical frequency effects arise will differ between a typical and atypical form of orthography, and looks to a target word will also differ between the two forms of words. In order to control for a considerable difference in visual features between logographic and syllabic scripts, e.g., logographic script ("樹", /ki/, 'tree') vs. syllabic script（"き", /ki/, 'tree'), we will use a character stroke count as a proxy of visual complexity and include it as a control variable (Miwa, Libben & Ikemoto, 2016).

Second, since we were interested in three-way interactions and statistically controlled confound factors, our GAMM models were large and complex, which limited us for direct comparisons of a few variables of interests, such as a task type (Go-NoGo task vs. delayed naming task) and population (L1 vs. L2 listeners). Now that we have a general idea of the effects of reduction and P-O consistency, as well as their interactions, for L1 and L2 listeners, we can design a series of experiments to assess the effect of reduction and P-O consistency separately to directly compare differences between task types and listeners. For instance, our results appear to suggest that P-O consistency effects are stronger in L2 than in L1 listeners, and the processing cost of reduced forms is greater for L1 than for L2 listeners. We can assess these effects one by one so that we can reduce the number of predictors and maintain the statistical power of the models.

Finally, we need to examine the recordings of participants' namings during the delayed naming task. Research shows that P-O consistency influences the production of reduced forms (Bürki et al., 2012). In addition, it is also interesting to compare acoustic properties of our experimental items and participants' productions in both conditions (naming a reduced or unreduced forms) to explore how participants adjust their production of words as they hear either reduced or unreduced word forms, and compare the resulting acoustic properties of their namings

to what we found in Study 1 to explore how comparable (or incomparable) they are across different speech styles.

5.4 Conclusion

Although speech is highly variable, speakers and listeners seem to communicate with ease. In order to further our understanding of the underlying mechanisms of speech communication, we conducted three studies to investigate (1) how production of speech varies across different styles of speech, (2) how listeners interact with such variability, and (3) how visual language information interacts with auditory language information during spoken language comprehension. We found a high degree of phonetic variability and phonetically reduced pronunciations of words, particularly in casual speech. These acoustically reduced ambiguous words caused listeners to incur additional processing costs for comprehension. Our results show that listeners exert orthographic knowledge to increase efficacy in the processing of such words. More specifically, how the orthographic information interacts with the realization of reduced word forms differs as follows. First, it depends on the type of listeners (L1 and L2). While the processing cost of reduced forms was attenuated in inconsistent words in L2, the cost was attenuated in consistent words in L1. Second, for L2 listeners, it varies depending on the proficiency level of the listeners (advanced vs. beginner) and the task that listeners perform (Go-NoGo task vs. delayed naming task). The high proficiency learners showed the clear P-O consistency effect that weakens as L2 proficiency decreases in the delayed naming task, but in the Go-NoGo task, the basic proficiency learners exhibited the clear consistency effect and it weakened as L2 proficiency increase.

Bibliography

Adda-Decker, M. & Lamel, L. (2018). 4. discovering speech reductions across speaking styles and languages. In *Rethinking Reduction* (pp. 101 – 28). Berlin, Boston: De Gruyter Mouton.

Akita, Y. & Kawahara, T. (2005). Generalized statistical modeling of pronunciation variations using variable-length phone context. In *Proceedings. (ICASSP '05). IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005.*, volume 1, (pp. I/689–I/692 Vol. 1).

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.

Aylett, M. & Turk, A. (2004). The smooth signal redundancy hypothesis: a functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Language and speech*, *47*(Pt 1), 31–56.

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.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412.

Barry, W. & Andreeva, B. (2001). Cross-language similarities and differences in spontaneous speech patterns. *Journal of the International Phonetic Association*, *31*(1), 51–66.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48.

Bates, E. & Liu, H. (1996). Cued Shadowing. *Language and Cognitive Processes*, *11*(6), 577–582.

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

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.

Boersma, P. & Weenink, D. (2016). Praat: doing phonetics by computer [Computer program].

Brouwer, S., Mitterer, H., & Huettig, F. (2013). Discourse context and the recognition of reduced and canonical spoken words. *Applied Psycholinguistics*, *34*(3), 519–539.

Browman, C. P. & Goldstein, L. (1990). Tiers in articulatory phonology, with some implications for casual speech. In J. Kingston & M. E. Beckman (Eds.), *Papers in Laboratory Phonology*, volume 1 of *Papers in Laboratory Phonology* (pp. 341–376). Cambridge University Press.

Browman, C. P. & Goldstein, L. (1992). Articulatory Phonology: An Overview. *Phonetica*, *49*, 155–180.

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.

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.

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.

Bybee, Joan, a. (2007). *The Phonology of the Lexicon : Evidence from Lexical Diffusion.* Oxford University Press.

Cangemi, F. & Niebuhr, O. (2018). 9. rethinking reduction and canonical forms. In *Rethinking Reduction* (pp. 277 – 302). Berlin, Boston: De Gruyter Mouton.

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.

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.

Cohen, L. & Dehaene, S. (2004). Specialization within the ventral stream: The case for the visual word form area. *NeuroImage*, *22*(1), 466–476.

Coltheart, M., Curtis, B., Atkins, P., & Haller, M. (1993). Models of Reading Aloud: Dual-Route and Parallel-Distributed-Processing Approaches. *Psychological Review*, *100*, 589–608.

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review*, *108*, 204–256.

Connine, C. M. (2004). It's not what you hear but how often you hear it: On the neglected role of phonological variant frequency in auditory word recognition. *Psychonomic Bulletin and Review*, *11*(6), 1084–1089.

Connine, C. M., Ranbom, L. J., & Patterson, D. J. (2008). Processing variant forms in spoken word recognition: The role of variant frequency. *Perception & Psychophysics*, *70*(3), 403–411.

Cutler, A. (1998). The recognition of spoken words with variable representations. In *Proceedings of the ESCA workshop on sound patterns of spontaneous speech*, (pp. 83–92).

Cutler, A. & Otake, T. (1998). Assimilation of place in Japanese and Dutch. In *Proceedings of the Fifth International Conference on Spoken Language Processing*, volume 5, (pp. 1751–1754).

Cutler, A., Treiman, R., & van Ooijen, B. (2010). Strategic deployment of orthographic knowledge in phoneme detection. *Language and Speech*, *53*(3), 307–320.

Davidson, L. (2011). Characteristics of stop releases in American English spontaneous speech. *Speech Communication*, *53*(8), 1042–1058.

Diependaele, K., Ziegler, J. C., & Grainger, J. (2010). Fast phonology and the bimodal interactive activation model. *European Journal of Cognitive Psychology*, *22*(5), 764–778.

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.

Dilts, P. C. (2013). *Modelling phonetic reduction in a corpus of spoken English using Random Forests and Mixed-Effects Regression*. PhD thesis, University of Alberta.

Dornbusch, T. (2012). *Orthographic influences on L2 auditory word processing*. PhD thesis, Technische Universität Dortmund.

Duffau, H. (2008). The anatomo-functional connectivity of language revisited. New insights provided by electrostimulation and tractography. *Neuropsychologia*, *46*(4), 927–934.

Ernestus, M. (2000). *Voice assimilation and segment reduction in casual Dutch a corpus-based study of the phonology-phonetics interface*. Utrecht: LOT.

Ernestus, M., Baayen, H., & Schreuder, R. (2002). The recognition of reduced word forms. *Brain and language*, *81*(1-3), 162–173.

Ernestus, M. & Warner, N. (2011). An introduction to reduced pronunciation variants. *Journal of Phonetics*, *39*(3), 253–260.

Escudero, P., Hayes-harb, R., & Mitterer, H. (2008). Novel second-language words and asymmetric lexical access. *Journal of Phonetics*, *36*(2), 345–360.

Escudero, P., Simon, E., & Mulak, K. E. (2014). Learning words in a new language: Orthography doesn't always help. *Bilingualism*, *17*(2), 384–395.

Escudero, P. & Wanrooij, K. (2010). The effect of L1 orthography on non-native vowel perception. *Language and Speech*, *53*(3), 343–365.

Forster, K. I. & Chambers, S. M. (1973). Lexical access and naming time. *Journal of Verbal Learning and Verbal Behavior*, *12*(6), 627 – 635.

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), 77–91.

Frost, R., Katz, L., & Bentin, S. (1987). Strategies for Visual Word Recognition and Orthographical Depth: A Multilingual Comparison. *Journal of Experimental Psychology: Human Perception and Performance*, *13*(1), 104–115.

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 & Cognition*, *44*(4), 554–564.

Godfrey, J. J., Holliman, E. C., & McDaniel, J. (1992). Switchboard: telephone speech corpus for research and development. In *[Proceedings] ICASSP-92: 1992 IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 1, (pp. 517–520 vol.1).

Goldinger, S. D. & Papesh, M. H. (2012). Pupil Dilation Reflects the Creation and Retrieval of Memories. *Current Directions in Psychological Science*, *21*, 90–95.

Grainger, J. & Ferrand, L. (1996). Masked orthographic and phonological priming in visual word recognition and naming: Cross-task comparisons. *Journal of Memory and Language*, *35*(5), 623–647.

Grainger, J., Spinelli, E., Farioli, F., Diependaele, K., & Ferrand, L. (2003). Masked Repetition and Phonological Priming within and across Modalities. *Journal of Experimental Psychology: Learning Memory and Cognition*, *29*(6), 1256–1269.

Greenberg, S. (1999). Speaking in shorthand - a syllable-centric perspective for understanding pronunciation variation. *Speech Communication*, *29*(2), 159–176.

Hannagan, T., Magnuson, J. S., & Grainger, J. (2013). Spoken word recognition without a TRACE. *Frontiers in Psychology*, *4*(SEP), 1–17.

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

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.

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.

Haynes, R. M., White, L., & Mattys, S. L. (2015). What do we expect spontaneous speech to sound like? In *18th International Congress of Phonetic Sciences, ICPhS 2015, Glasgow, UK, August 10-14, 2015*.

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 Human perception and performance*, *43*(1), 126–146.

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), 1110–51.

Hubert, I. & Järvikivi, J. (2019). Dark forces in language comprehension: The case of neuroticism and disgust in a pupillometry study. In Goel, A. K., Seifert, C. M., & Freksa, C. (Eds.), *Annual Meeting of the Cognitive Science Society. Montreal (CogSci 2019)*, (pp. 450–456)., Montreal, QC. Proceedings of the 41st Annual Conference of the Cognitive Science Society.

Itou, K., Yamamoto, M., Takeda, K., Takezawa, T., Matsuoka, T., Kobayashi, T., Shikano, K., & Itahashi, S. (1999). Jnas: Japanese speech corpus for large vocabulary continuous speech recognition research. *Journal of the Acoustical Society of Japan (E)*, *20*(3), 199–206.

Jakimik, J., Cole, R. A., & Rudnicky, A. I. (1985). Sound and spelling in spoken word recognition. *Journal of Memory and Language*, *24*(2), 165 – 178.

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.

Johnson, K. (2004). Massive reduction in conversational American English. In *Spontaneous speech: Data and analysis. Proceedings of the 1st session of the 10th international symposium*, (pp. 29–54). Tokyo, Japan: The National International Institute for Japanese Language.

Kahneman, D. & Beatty, J. (1966). Pupil Diameter and Load on Memory. *Science*, *154*(3756), 1583–1585.

Katz, L. & Frost, R. (1992). Chapter 4 the reading process is different for different orthographies: The orthographic depth hypothesis. In R. Frost & L. Katz (Eds.), *Orthography, Phonology, Morphology, and Meaning*, volume 94 of *Advances in Psychology* (pp. 67 – 84). North-Holland.

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.

Kawahara, S. (2017). Durational compensation within a CV mora in spontaneous Japanese: Evidence from the Corpus of Spontaneous Japanese. *The Journal of the Acoustical Society of America*, *142*(EL143), 1–6.

Keune, K., Ernestus, M., van Hout, R., & Baayen, R. H. (2005). Variation in dutch: From written mogelijk to spoken mok. *Corpus Linguistics and Linguistic Theory*, *1*(2), 183 – 223.

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

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.

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.

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.

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

Laan, G. P. (1997). The contribution of intonation, segmental durations, and spectral features to the perception of a spontaneous and a read speaking style. *Speech Communication*, *22*(1), 43 – 65.

Labov, W. (1972). *Sociolinguistic patterns*. University of Pennsylvania Press.

Laeng, B., Sirois, S., & Gredebäck, G. (2012). Pupillometry: A Window to the Preconscious? *Perspectives on Psychological Science*, *7*(1), 18–27.

Landerl, K. (2005). Reading acquisition in different orthographies: Evidence from direct comparisons. In R. M. Joshi & P. G. Aaron (Eds.), *Handbook of Orthography and Literacy* (pp. 629–650). Mahwah, NJ: Lawrence Erlbaum Associates.

Lenth, R. (2019). *emmeans: Estimated Marginal Means, aka Least-Squares Means*. R package version 1.3.3.

Lindblom, B. (1990). Explaining phonetic variation: A sketch of the h&h theory. In W. J. Hardcastle & A. Marchal (Eds.), *Speech Production and Speech Modelling* (pp. 403–439). Dordrecht: Springer Netherlands.

Lõo, K., Järvikivi, J., Tomaschek, F., Tucker, B. V., & Baayen, R. H. (2018). Production of Estonian case-inflected nouns shows whole-word frequency and paradigmatic effects. *Morphology*, *28*.

Maekawa, K. (2003). Corpus of Spontaneous Japanese: Its design and evaluation. In *Proceedings of the ISCA and IEEE Workshop on Spontaneous Speech Processing and Recognition (SSPR2003), Tokyo*, (pp. 7–12).

Maekawa, K. (2005). Toward a pronunciation dictionary of Japanese: Analysis of CSJ. In *Proceedings of Symposium on Large-Scale Knowledge Resources (LKR2005)*, (pp. 43–48).

Maekawa, K. (2015). 16 Corpus-based phonetics. In *Handbook of Japanese Phonetics and Phonology* (pp. 651 – 680). Berlin, Boston: De Gruyter Mouton.

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.

Manuel, S. Y., Shattuck-Hufnagel, S., Huffman, M. K., Stevens, K. N., Carlson, R., & Hunnicutt, S. (1992). Studies of Vowel and Consonant Reduction. In *Second conference on spoken language processing*, (pp. 943–946)., Banff, Alberta.

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.

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.

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.

Maurer, U., Zevin, J. D., & Mccandliss, B. D. (2008). Left-lateralized N170 effects of visual expertise in reading: evidence from Japanese syllabic and logographic scripts. *Journal of Cognitive Neuroscience*, *20*(10), 1878–1891.

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.

McGurk, H. & Macdonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264*, 746–748.

McQueen, J. (1996). Word Spotting. *Language and Cognitive Processes*, *11*(6), 695–699.

Mehta, G. & Cutler, A. (1988). Detection of target phonemes in spontaneous and read speech. *Language and Speech*, *31*(2), 135–156.

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.

Mitsugi, S. (2018). Proficiency influences orthographic activations during l2 spoken-word recognition. *International Journal of Bilingualism*, *22*(2), 199–214.

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.

Miwa, K., Libben, G., & Ikemoto, Y. (2016). Visual trimorphemic compound recognition in a morphographic script. *Language, Cognition and Neuroscience*, *3798*(August).

Mizoguchi, A. (2019). *Articulation of the Japanese moraic nasal: Place of articulation, assimilation, and L2 transfer*. phdthesis, City University of New York.

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 Dmyterko, E. (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*.

Muneaux, M. & Ziegler, J. C. (2004). Locus of orthographic effects in spoken word recognition: Novel insights from the neighbour generation task. *Language and Cognitive Processes*, *19*(5), 641–660.

Muthusamy, Y. K., Cole, R. a., & Oshika, B. T. (1992). The OGI Multi-Language Telephone Speech Corpus. In *International Conference on Spoken Language Processing 92*, (pp. 895–898)., Banff, Alberta.

Nakamura, M., Iwano, K., & Furui, S. (2008). Differences between acoustic characteristics of spontaneous and read speech and their effects on speech recognition performance. *Computer Speech and Language*, *22*(2), 171–184.

National Language Research Institute (1970). *Denshi-keisanki niyoru shinbun no goichousa [Studies on the vocabulary of modern newspapers (Vol. 1), General descriptions and vocabulary frequency tables]*. Tokyo, Japan: Shuei Shuppan.

National Language Research Institute (1993). *Bunrui-goi hyou (furoppi ban) [Thesaurus (Floppy Disk Version)]*. Tokyo, Japan: Shuei Shuppan.

Neef, M. & Balestra, M. (2011). Measuring graphematic transparency: German and Italian compared. *Written Language & LiteracyWritten Language and Literacy*, *14*(1), 109–142.

Papesh, M. H. & Goldinger, S. D. (2012). Pupil-BLAH-metry: Cognitive effort in speech planning reflected by pupil dilation. *Attention, Perception, & Psychophysics*, *74*(4), 754–765.

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., Kolinsky, R., Ventura, P., Radeau, M., & Morais, J. (2007). Orthographic representations in spoken word priming: no early automatic activation. *Language and speech*, *50*(Pt 4), 505–31.

Pattamadilok, C., Morais, J., Colin, C., & Kolinsky, R. (2014). Unattentive speech processing is influenced by orthographic knowledge: Evidence from mismatch negativity. *Brain and Language*, *137*, 103–111.

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.

Perre, L., Bertrand, D., & Ziegler, J. C. (2011). Literacy affects spoken language in a non-linguistic task: An ERP study. *Frontiers in Psychology*, *2*(OCT), 1–8.

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.

Perry, C., Ziegler, J. C., & Zorzi, M. (2007). Nested incremental modeling in the development of computational theories: The CDP+ model of reading aloud. *Psychological Review*, *114*(2), 273–315.

Perry, C., Ziegler, J. C., & Zorzi, M. (2010). Beyond single syllables: Large-scale modeling of reading aloud with the Connectionist Dual Process (CDP++) model. *Cognitive Psychology*, *61*(2), 106–151.

Pitt, M. A., Dilley, L., Johnson, K., Kiesling, S., Raymond, W., Hume, E., & Fosler-Lussier, E. (2007). Buckeye corpus of conversational speech (2nd release)[www. buckeyecorpus. osu. edu] columbus, OH: Department of psychology.

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.

Porretta, V. & Tucker, B. V. (2019). Eyes Wide Open: Pupillary Response to a Foreign Accent Varying in Intelligibility. *Frontiers in Communication*, *4*(February), 1–12.

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.

Pylkkänen, L. & Okano, K. (2010). The nature of abstract orthographic codes: Evidence from masked priming and magnetoencephalography. *PLoS ONE*, *5*(5).

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.

Qu, Q. & Damian, M. F. (2017). Orthographic effects in spoken word recognition: Evidence from Chinese. *Psychonomic Bulletin and Review*, *24*(3), 901–906.

R Development Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

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.

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.

Ranbom, L. J. & Connine, C. M. (2011). Silent letters are activated in spoken word recognition. *Language and Cognitive Processes*, *26*(2), 236–261.

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), 1588–94.

Rumelhart, D. E. & McClelland, J. L. (1982). An interactive activation model of context effects in letter perception: Part 2. The contextual enhancement effect and some tests and extensions of the model. *Psychological Review*, *89*(1), 60–94.

Saito, H., Masuda, H., & Kawakami, M. (1998). Form and sound similarity effects in kanji recognition. *Reading and Writing: An Interdisciplinary Journal*, *10*, 323–357.

Salverda, A. P. & Tanenhaus, M. K. (2010). Tracking the time course of orthographic information in spoken-word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(5), 1108–1117.

Seidenberg, M. S. & Tanenhaus, M. K. (1979). Orthographic effects on rhyme monitoring. *Journal of Experimental Psychology Human Learning and Memory*, *5*(6), 546–554.

Seymour, P. H., Aro, M., & Erskine, J. M. (2003). Foundation literacy acquisition in European orthographies. *British Journal of Psychology*, *94*(2), 143–174.

Shockey, L. (2003). *Sound patterns of spoken English*. Oxford: Blackwell Publishing.

Sóskuthy, M. (2017). Generalised additive mixed models for dynamic analysis in linguistics: a practical introduction. *arXiv e-prints*, arXiv:1703.05339.

Stone, G., 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.

Taft, M. (2006). Orthographically influenced abstract phonological representation: Evidence from Non-rhotic speakers. *Journal of Psycholinguistic Research*, *32*(1), 67–78.

Taft, M. (2011). Orthographic influences when processing spoken pseudowords: Theoretical implications. *Frontiers in Psychology*, *2*, 1–7.

Titone, C. M. C. D. (1996). Phoneme Monitoring. *Language and Cognitive Processes*, *11*(6), 635–646.

Torreira, F. & Ernestus, M. (2011). Realization of voiceless stops and vowels in conversational french and spanish. *Laboratory Phonology*, *2*(2), 331–353.

Tucker, B. V. (2007). *Spoken word recognition of the reduced American English flap*. PhD 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.

Tucker, B. V. & Ernestus, M. (2016). Why we need to investigate casual speech to truly understand language production, processing and the mental lexicon. *The Mental Lexicon*, *11*(3), 375–400.

Turnbull, R. (2018). Patterns of probabilistic segment deletion/reduction in English and Japanese. *Linguistics Vanguard*, *4*(2s).

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.

van Dommelen, W. A. (2018). 3. reduction in native and non-native read and spontaneous speech. In F. Cangemi, M. Clayards, O. Niebuhr, B. Schuppler, & M. Zellers (Eds.), *Rethinking Reduction* (pp. 73 – 100). Berlin, Boston: De Gruyter Mouton.

van Rij, J. (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, 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.

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.

Van Son, R. & Pols, L. C. (1999). An acoustic description of consonant reduction. *Speech communication*, *28*(2), 125–140.

Vance, T. J. (1987). *An introduction to Japanese phonology*. New York: University of New York Press.

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.

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*(July).

Veivo, O., Porretta, V., Hyönä, J., & Järvikivi, J. (2018). Spoken second language words activate native language orthographic information in late second language learners. *Applied Psycholinguistics*, *1*(22), 1–22.

Ventura, P., Morais, J., & Kolinsky, R. (2007). The development of the orthographic consistency effect in speech recognition: From sublexical to lexical involvement. *Cognition*, *105*(3), 547–576.

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.

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.

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. PMID: 30362403.

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* chapter 8, (pp. 185–204). Psychology Press.

Warner, N. (2011). Reduction. In M. van Oostendorp, C. J. Ewen, E. Hume, & K. Rice (Eds.), *The Blackwell companion to phonology* chapter 79, (pp. 1866–1891). Oxford: Oxford University Press.

Warner, N. (2012). The oxford handbook of laboratory phonology. In M. K. H. Abigail C. Cohn, Cécile Fougeron (Ed.), *Methods for studying spontaneous speech* (pp. 621–633). Oxford University 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.

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.

Wieling, M., Tomaschek, F., Arnold, D., Tiede, M., Bröker, F., Thiele, S., Wood, S. N., & Baayen, R. H. (2016). Investigating dialectal differences using articulography. *Journal of Phonetics*, *59*, 122–143.

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.

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

Wood, S. N. (2006). *Generalized additive models: An introduction with R*. Boca Raton, FL: Chapman & Hall/CRC Press.

Wydell, T. N. (1998). What matters in kanji word naming: Consistency, regularity, or On/Kun-reading difference? *Reading and Writing*, *10*, 359–373.

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.

Wydell, T. N., Patterson, K. E., & Humphreys, G. W. (1993). Phonologically mediated access to meaning for Kanji: Is a rows still a rose in Japanese Kanji? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*(3), 491–514.

Yoneyama, K. (2002). *Phonological neighborhoods and phonetic similarity*. PhD 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.

Zekveld, A. A., Koelewijn, T., & Kramer, S. E. (2018). The Pupil Dilation Response to Auditory Stimuli: Current State of Knowledge. *Trends in Hearing*, *22*, 233121651877717.

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

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

Zellers, M., Schuppler, B., & Clayards, M. (2018). 1. introduction, or: why rethink reduction? In *Rethinking Reduction* (pp. 1 – 24). Berlin, Boston: De Gruyter Mouton.

Ziegler, J. C. & Ferrand, L. (1998). Orthography shapes the perception of speech: The consistency effect in auditory word recognition. *Psychonomic Bulletin & Review*, *5*(4), 683–689.

Ziegler, J. C., Ferrand, L., & Montant, M. (2004). Visual phonology: the effects of orthographic consistency on different auditory word recognition tasks. *Memory & cognition*, *32*(5), 732–741.

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.

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.

Ziegler, J. C., Van Orden, G. C., & Jacobs, a. M. (1997). Phonology can help or hurt the perception of print. *Journal of experimental psychology. Human perception and performance*, *23*(3), 845–860.