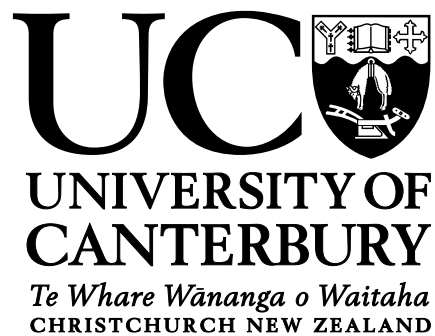


# Second-Language Acquisition of a Sublexicon Phonology: Loanword Phonology and Phonotactics in Japanese

Wakayo Mattingley

This dissertation is submitted in partial fulfilment  
of the requirements for the degree of  
Doctor of Philosophy in Linguistics



Department of Linguistics  
University of Canterbury  
New Zealand  
September 2020

# Abstract

Language users and learners are sensitive to distributional information in their environment, which enables them to extract regularities that occur in the language input that they are exposed to. This process is referred to as statistical learning. While the statistical learning phonotactic literature thoroughly investigates the learning of overall phonotactics in specific languages, little is known about cases where different phonological systems coexist within a single language. The Japanese lexicon is generally classified into four lexical strata according to the etymological status of each word (Itô & Mester, 1995, 1999, 2001). Although each stratum includes the internal phonological similarity in the Japanese language as a whole, there are also distinctive phonological properties. A recent study suggests that language users should be able to learn phonotactics of each sublexicon based on the same kind of statistical probabilities that computers analyse from language users' accumulated lexicons (Morita, 2018). This thesis examines whether second-language (L2) learners can learn the loanword phonotactics/phonology of Japanese through experience of using and/or passive exposure to Japanese lexical stratification. Using two loanword phonological regularities (categorical and gradient rules) as a case study, two fully-crossed perceptual experiments involving English-speaking learners of Japanese, native speakers of Japanese, and English-speaking monolinguals are presented.

The first experiment explores listeners' phonotactic/phonological knowledge of nativised loanwords in Japanese using a well-formedness task which shows the adaptation of English final consonants in monosyllabic words. Listeners judge whether the pronunciation they hear is how the word would be pronounced if it was a Japanese word, rating how confident they are on a scale of 1-5. This study shows that L2 learners learn categorical rules, but not gradient patterns. This study also confirms that loanword phonotactics and overall phonotactics make separate contributions to perceived well-formedness. L2 learners access and make use of the sublexicon-specific probabilities of Japanese during the task. The second perceptual experiment is designed to support the findings in the first experiment, by testing for discrimination of non-native consonantal contrasts. Even under high memory demand, L2 learners show the ability to discriminate non-native consonantal contrasts (i.e., CVCV/CVCCV) effectively enough to support findings in the first experiment.

These results suggest that L2 learners can implicitly detect the statistical structure of a language's sublexicon phonology over the course of acquiring a natural language. However,

while native speakers of Japanese learn a gradient rule, L2 learners of Japanese do not. A potential explanation for the differences in gradient rule learning is that the vocabulary size of the target language might play a crucial role. This remains an open question.

In addition, the present work provides a basis for future investigation into whether L2 learners of Japanese, whose native language is other than English, are able to learn Japanese loanword phonotactics/phonology. L1 English-L2 Japanese speakers might gain advantage in perceiving the English input which inevitably overlaps with the phonological form of the host language.

# Acknowledgements

I am greatly thankful for everyone who helped me out in some way, shape or form throughout the path to my PhD. Without those people who supported and encouraged me during the PhD journey, I could not have this day. First and foremost, I would like to sincerely and gratefully thank my supervisory team: Kevin Watson, Jen Hay, and Shigeto Kawahara. To Kevin, I would like to thank you for your constant support, endless encouragement and providing me constructive feedback. Your valuable suggestions and great jokes during our meetings always made me relaxed and invigorated. Jen, thank you for keeping me on track throughout the project. Your endless support, patience, providing me insightful comments and suggestions and how you introduced me to various research methods helped me learn so much during my doctoral study. To my associative supervisor Shigeto, thank you for giving me immense support, invaluable feedback and guiding me to important research studies at any time despite being in another country. I cannot thank you enough for the time and energy that you spent for me. I feel privileged to have been able to learn so much about being an independent researcher from each of you. I also thank my two external examiners Stefan Frisch and Hunter Hatfield for their insightful feedback and guiding me to important literature.

I would like to thank Beth Hume for all the opportunities you gave me, believing in my potential to pursue this academic journey. Without you, I definitely could not enjoy this sense of accomplishment. Kathleen Currie Hall, thank you for teaching me many new valuable skills to be a researcher. I would also like to extend sincere thanks to Viktoria Papp for developing my knowledge of R and Praat, as well as our many good conversations and your advice along the way. I also thank all the faculty, visitors and students at the University of Canterbury and the NZILBB (New Zealand Institute of Language, Brain and Behaviour). I have benefited so much from socio-meetings, stats talks, seminars, syntax sessions, and social events. Especially, I am grateful to James Brand, Jeremy Needle, and Yoonmi Oh for your huge support in helping to run my two experiments and data analysis. Mineko Shirakawa, thank you for being a great *senpai* and amazing speaker, who helped me to create the stimuli used in the experiments. Daiki Hashimoto, thank you for the friendship, sharing with me your great knowledge of Japanese phonology, and insightful discussion about anything, anytime. I also deeply appreciate great support from Clay Beckner, Lynn Clark, Donald Derrick, Jonathan Dunn, Robert Fromont, Susan Foster-Cohen, Maria Hellstrom, Sarah Hawkins, Emma Parnell, Jeanette King, Heidi Quinn, Petya Ràcz, Dineke Schokkin, Simon Todd and Dan Villarreal. I would also express my

gratitude to the PhD students for being supportive and fantastic friends: Moonsun Choi, Marie Fournier, Andy Gibson, Mohammed Gwasmeh, Jacq Jones, Stephanie Kaefer and Muneir Mohammed.

There are so many faculty members who supported me during the recruitment of participants, thus brightening me up during the hard times in my PhD study. I am grateful for the enormous support of the Department of Japanese at the University of Canterbury: Masayoshi Ogino, Susan Bouterey, Eri Kojima-Mathieson, and Henk de Groot. Thank you for allowing me to recruit participants before each class starting and encouraging students to participate in my study. I also received a generous support from faculty members at other New Zealand universities: Harumi Minagawa and Mori Michiyo, at the University of Auckland, and Junji Kawai at the Auckland University of Technology. I am grateful to CCEL who helped me by allowing me to recruit from among their students. I wish to sincerely thank all who participated in my research for your cooperation, as well as those who introduced me to the participants. Without their voluntary participation, this study could not have been done.

I would also express my gratitude to Karen McLean who proofread this thesis, also providing me many useful tips for writing which enhanced the clarity of my work. To my friends inside and outside of New Zealand, thank you all for supporting my pilot study, recruiting participants, and giving me a relaxing time and great food that took my mind off the hard work when I needed a break.

I am also fortunate to have had financial support throughout my PhD. This research was founded by the University of Canterbury's doctoral scholarship and research grants from LSAP funding.

I would like to thank my parents for their constant encouragement. Finally, to my husband, Steven Mattingley, thank you for your love, understanding, patience, support and taking care of me. During my longstanding PhD journey, I sometimes had to pass over bumpy and crooked roads, but you always turn on the light on my way to get here today. I could not come so far without you.

# Contents

<b>Chapter 1 Introduction and Literature Review .....</b>	<b>1</b>
1.1 Introduction.....	1
1.2 Statistical Learning and Acquisition of Phonotactics .....	7
1.2.1 The effects of statistical knowledge on word recognition, production, and perception in L1s .....	7
1.2.2 The effects of vocabulary size on statistical learning .....	9
1.3 Rule Learning in Artificial Languages .....	10
1.4 Acquisition of L2 Phonotactic Knowledge.....	14
<b>Chapter 2 Lexical Phonology of Japanese.....</b>	<b>20</b>
2.1 Lexical Stratification of the Japanese Language .....	20
2.2 Japanese Phonology .....	22
2.2.1 Phoneme inventory.....	22
2.2.2 Syllable structure.....	25
2.2.3 Psychological reality of lexical strata in Japanese .....	26
2.3 Loanword Phonology.....	29
2.3.1 Introduction .....	29
2.3.2 Loanwords in Japanese.....	31
2.3.2.1 Vowel epenthesis .....	31
2.3.2.2 Consonant gemination in loanwords.....	35
2.3.2.3 Corpus work for consonant gemination.....	38
2.3.3 Frequency of loanwords and estimated vocabulary size of loanwords.....	40
2.3.3.1 Vocabulary size of Japanese .....	40
2.3.3.2 Frequency of loanwords and estimated vocabulary size of loanwords.....	41
2.4 Research Questions and Predictions .....	42
<b>Chapter 3 Pilot Study: Confidence-Rating Task for Adaptation of English Word-Final Consonants.....</b>	<b>45</b>
3.1 Introduction.....	45
3.2 Research Design and Methodology .....	46
3.2.1 Materials.....	46
3.2.2 Audio stimuli.....	50
3.2.3 Acoustic characteristics of the audio stimuli.....	51
3.2.4 Participants .....	55
3.2.5 Procedure.....	55

3.3	Overall Outcome and Application to Actual Experiment.....	56
3.3.1	Duration of the task.....	56
3.3.2	Rating.....	56
3.3.2.1	CVC tokens.....	56
3.3.2.2	Interpretation of “correct”.....	58
3.4	Summary.....	58
<b>Chapter 4 Confidence-Rating Task: Adaptation of English Word-Final Consonants.....</b>		<b>60</b>
4.1	Research Design and Methodology.....	61
4.1.1	Participants.....	61
4.1.2	Materials.....	62
4.1.2.1	Stimuli.....	62
4.1.2.2	Extracting for statistical phonotactic scores of Japanese.....	62
4.1.2.3	Phonotactic probability of stimuli.....	65
4.1.3	Procedure.....	68
4.1.4	Questionnaires.....	68
4.1.5	Principal component analysis (PCA).....	69
4.1.5.1	PCA of native Japanese speakers.....	70
4.1.5.2	PCA of NZE-speaking learners of Japanese.....	72
4.1.5.3	PCA of NZE-speaking non-learners of Japanese.....	74
4.2	Phonotactic Effects.....	76
4.2.1	Statistical analysis.....	76
4.2.2	Results.....	76
4.3	Statistical Analysis for Phonological Processing.....	80
4.3.1	Factors.....	80
4.3.1.1	Phonological factors.....	81
4.3.1.2	Extra-linguistic factors.....	81
4.3.2	Mixed effects models for phonological process (i.e., pronunciation type).....	82
4.4	Results.....	83
4.4.1	Phonological process: Full dataset.....	83
4.4.2	Phonological process: Differences between native Japanese speakers and learners of Japanese.....	86
4.4.3	Phonological process: Differences between learners and non-learners of Japanese.....	88
4.4.4	Phonological process: Native Japanese speakers.....	92
4.4.4.1	Voiced geminates relation to POA.....	95
4.4.4.2	Rating for CVC.....	96

4.4.5	Phonological process: Learners of Japanese .....	99
4.4.5.1	Effects of phonological factors .....	101
4.4.5.2	Effects of extra-linguistic factors .....	104
4.4.6	Phonological process: Non-learners of Japanese .....	105
4.4.6.1	Effects of phonological factors .....	106
4.4.6.2	Effects of extra-linguistic factors .....	108
4.5	Discussion .....	109
4.5.1	Summary of results.....	110
4.5.1.1	Categorical rules: Quality of epenthetic vowels .....	110
4.5.1.2	Gradient phonological rules: Consonant gemination.....	111
4.5.1.3	The effect of exposure to the target language (PC-JJ, PC-JNZ).....	112
4.5.1.4	Rating for CVC .....	114
4.5.2	Discussion .....	114
4.5.3	Remaining issues.....	117
4.6	Conclusion .....	118

**Chapter 5 Auditory Memory Decision Task: Discriminating Non-native Segmental Contrasts in Spoken Word Lists.....120**

5.1	Introduction.....	120
5.2	Background for Memory Decision Task.....	122
5.2.1	The effects of attention control for discrimination of non-native length contrasts. ....	122
5.2.2	Task demand, memory load, and speech processing .....	124
5.2.3	The role of phonetic salience on non-native perception of contrast .....	126
5.2.4	Current study .....	128
5.2.5	Predictions.....	129
5.3	Research Design and Methodology .....	132
5.3.1	Materials.....	132
5.3.2	Audio stimuli.....	133
5.3.3	Acoustic characteristics of the audio stimuli.....	134
5.3.4	Audio clips as stimulus lists .....	139
5.3.5	Auditory memory decision task protocol .....	141
5.3.6	Participants .....	142
5.3.7	Procedure.....	143
5.3.8	Questionnaires and PCA .....	145
5.3.8.1	Questionnaires.....	145
5.3.8.2	Principal component analysis (PCA).....	145



5.3.9	Statistical analysis .....	147
5.3.9.1	Factors.....	147
5.3.9.2	Mixed effects models.....	149
5.4	Results.....	150
5.4.1	Singleton-Geminate contrasts .....	150
5.4.1.1	Native Japanese listeners .....	151
5.4.1.2	L1 English-L2 Japanese listeners (learners of Japanese).....	152
5.4.1.3	Monolingual English listeners .....	155
5.4.2	Results of Singleton-CVC contrasts.....	158
5.4.2.1	Monolingual English listeners .....	158
5.4.2.2	L1 English-L2 Japanese listeners (learners of Japanese).....	159
5.4.2.3	Native Japanese listeners .....	162
5.5	Discussion .....	165
5.5.1	Summary of results.....	165
5.5.1.1	Singleton-Geminate contrasts .....	165
5.5.1.2	Singleton-CVC contrasts .....	168
5.5.2	Implication for the first experiment .....	170
5.5.2.1	Implication of singleton and gemination ratings by L2 learners .....	171
5.5.2.2	Implication of CVC rating by Japanese listeners.....	172
5.5.3	General implications .....	172
5.5.3.1	Effect of acoustic salience .....	172
5.5.3.2	Effects of L2 language experiences on L1 perception.....	174
5.5.3.3	Attention control .....	175
5.5.3.4	The importance of type of task .....	176
5.6	Conclusion .....	177
<b>Chapter 6 Discussion and Conclusion.....</b>		<b>179</b>
6.1	Summary of Research Question, Predictions and Findings.....	179
6.1.1	Summary of confidence-rating task (well-formedness task).....	179
6.1.2	Summary of auditory memory decision task.....	183
6.1.3	Bringing corpus study, well-formedness judgement and perceptual discrimination together.....	187
6.2	General discussion .....	191
6.2.1	Influence of L1 on L2 statistical learning: Lexical activation and priming .....	191
6.2.2	The effect of individual degree of exposure to L2 .....	193
6.2.3	From perceptual knowledge to higher level phonological knowledge .....	195
6.2.4	Influence of vocabulary size on statistical learning .....	196

6.2.5 Possible constraints for statistical learning .....	199
6.3 Conclusion .....	201
<b>References .....</b>	<b>203</b>
<b>Appendix A: List of items for pilot study .....</b>	<b>226</b>
<b>Appendix B: List of items for well-formedness judgment task .....</b>	<b>228</b>
<b>Appendix C: Preregistration document for AsPredicted .....</b>	<b>230</b>
<b>Appendix D: Questionnaires .....</b>	<b>232</b>
<b>Appendix E: Preregistration document for AsPredicted .....</b>	<b>240</b>
<b>Appendix F: Short lists for memory decision task.....</b>	<b>242</b>

# List of Figures

Figure 2.1 Lexical strata in a core-periphery structure adapted from Itô and Mester (1999)...	20
Figure 2.2 Frequency of geminates and singletons for source words with a word-final stop following a lax vowel in the BCCWJ. ....	39
Figure 2.3 Frequency of loanword in every 1000 overall vocabularies in the BCCWJ. ....	41
Figure 3.1 F1 and F2 values for individual vowels from the stimulus speaker. ....	52
Figure 3.2 Boxplots of the ratio of vowel-to-word durations for the stimulus speaker. ....	53
Figure 3.3 Boxplots of the ratio of closure-to-word duration for the stimulus speaker. ....	54
Figure 4.1 The correlation between the log-phonotactic scores for overall Japanese and that of loanwords in all stimulus words. ....	66
Figure 4.2 The correlation between the loanword log-transformed and overall log-transformed phonotactic probabilities for each place of articulations. ....	67
Figure 4.3 Scree plot from PCA of native Japanese speakers. ....	71
Figure 4.4 Scree plot from PCA of L2 learners of Japanese. ....	73
Figure 4.5 Scree plot from PCA of non-learners of Japanese. ....	75
Figure 4.6 The effects of phonotactic scores on responses for each group: voiceless stimuli. ....	78
Figure 4.7 The effects of phonotactic scores on responses for each group: voiced stimuli. ....	79
Figure 4.8 Plots of the three-way interaction between VOICING TYPE, PRONUNCIATION TYPE and GROUP in the full dataset. ....	85
Figure 4.9 Plots of the three-way interaction between VOICING TYPE and PRONUNCIATION TYPE and GROUP. ....	87
Figure 4.10 Plots of the two-way interaction between VOICING TYPE and PRONUNCIATION TYPE. ....	90
Figure 4.11 Plots of the two-way interaction between PRONUNCIATION TYPE and GROUP. ....	91
Figure 4.12 Plots of the three-way interaction between PRONUNCIATION TYPE, VOICING TYPE and POA. ....	94
Figure 4.13 Mean response ratings of individual native Japanese speaker for gemination minus non-gemination ratings in voiced context by place of articulation. ....	96
Figure 4.14 Plots of the three-way interaction between VOICING TYPE, POA, and PC-ENGLISH (exposure to English in NZ) in the native speakers of Japanese dataset. ....	98

Figure 4.15 Plots of the two-way interaction between POA and PC-Language (exposure to other languages). .....	99
Figure 4.16 Plots of the three-way interaction between VOICING TYPE, PRONUNCIATION TYPE and POA in the learners of Japanese dataset. ....	102
Figure 4.17 Plots of the three-way interaction between VOICING TYPE, PRONUNCIATION TYPE and PC-JNZ in the learners of Japanese dataset. ....	104
Figure 4.18 Plots of the two-way interaction between PRONUNCIATION TYPE and POA for non-learners of Japanese. ....	107
Figure 4.19 The plot of the interaction between PRONUNCIATION TYPE and PC-JNZ (exposure to Japanese in NZ).....	109
Figure 5.1 F1 and F2 values for individual vowels from the stimulus speaker. ....	135
Figure 5.2 Violin plots of ratio of closure duration of singleton and geminate stops for the stimulus speaker.....	137
Figure 5.3 Violin plots of ratio of closure (C) to word duration for the stimulus speaker: (a) single consonant (b) geminate consonant. ....	138
Figure 5.4 Violin plots of ratio of closure to word duration for the stimulus speaker: (a) single consonant (b) geminate consonant.....	138
Figure 5.5 Auditory memory decision task protocol. ....	142
Figure 5.6 Distribution and probability density of discrimination accuracy on singleton-geminate contrast by native Japanese listeners for the raw data (left) and predicted interaction between stimulus pair and PC-English (right).....	152
Figure 5.7 Distribution and probability density of discrimination accuracy on singleton-geminate contrast by L1 English-L2 Japanese listeners for the raw data (left) and predicted interaction between POA, voicing type and stimulus pair.....	154
Figure 5.8 Interaction between pair and PC-JJ (left), and between pair and PC-JNZ (right).155	
Figure 5.9 Distribution and probability density of discrimination accuracy on singleton-geminate contrast by monolingual English listeners for the raw data (left) and predicted interaction between voicing type and stimulus pair (right).....	157
Figure 5.10 Distribution and probability density of discrimination accuracy on singleton-CVC contrast by monolingual English listeners for the raw data (left) and predicted interaction between POA and stimulus pair.....	159
Figure 5.11 Discrimination and probability density of discrimination accuracy on singleton-CVC by L1 English-L2 Japanese listeners for the raw data. ....	161

Figure 5.12 Predicted interaction between POA, stimulus pair and PC-JJ (left), and interaction between pair and PC-JNZ (right).....	162
Figure 5.13 Distribution and probability density of discrimination accuracy on singleton–CVC contrast by native Japanese listeners for the raw data (left) predicted interaction between POA and stimulus pair (right). ....	164
Figure 5.14 Interaction between pair and PC-English. ....	164
Figure 6.1 Results from three different analysis regarding POA.....	188

# List of Tables

Table 2.1 Consonants of Japanese. Adapted from Akamatsu (2000) and Vance (2008) .....	22
Table 2.2 Basic correspondence between lax English vowels and their Japanese counterparts, Kubozono (2015, p. 315).....	32
Table 2.3 Basic correspondence of coda consonants, Kubozono (2015, p. 322) .....	33
Table 3.1 Examples for sets of stimuli and its structure: voiceless sets .....	48
Table 3.2 Examples for sets of stimuli and its structure: voiced sets .....	49
Table 3.3 Mean F1/F2/F3 formant values and standard deviations for voiced and voiceless contexts.....	51
Table 3.4 The number of token counts, the mean and SD in ms. of the five vowels along with vowel-to-word duration ratio.....	53
Table 3.5 Summary of amendment after pilot study.....	59
Table 4.1 Number of participants in each group for analysis .....	61
Table 4.2 Number of audio stimuli for each participant.....	62
Table 4.3 Two principal components revealed by PCA for native Japanese speakers.....	71
Table 4.4 Two principal components revealed by PCA for learners of Japanese .....	73
Table 4.5 One principal component revealed by PCA for non-learners of Japanese .....	75
Table 4.6 Factors considered in analysis of dataset.....	80
Table 4.7 Numbers of stimuli of each place of articulation by group .....	82
Table 4.8 Predictions for well-formedness .....	83
Table 4.9 Model summary for full dataset (all groups) .....	83
Table 4.10 Model summary for native Japanese speakers and learners of Japanese.....	86
Table 4.11 Comparison of predicted rating between voiceless and voiced contexts for each pronunciation type across Japanese group and learners' group .....	88
Table 4.12 Model summary for learners and non-learners of Japanese .....	89
Table 4.13 Model summary for native Japanese speakers.....	92
Table 4.14 Model summary for CVC .....	97
Table 4.15 Model summary for learners of Japanese .....	100
Table 4.16 Summary table for effects of pronunciation type on responses.....	103
Table 4.17 Model summary for non-learners of Japanese .....	106
Table 4.18 Summary comparison of predicted rating for each place of articulation.....	107
Table 5.1 Example of stimuli.....	133

Table 5.2 Mean F1/ F2/F3 formant values and standard deviations for voiced and voiceless contexts.....	134
Table 5.3 The number of token counts, the mean and SD in ms. of the five vowels .....	136
Table 5.4 Mean duration (ms) of word, stop closure and mean ratios of geminate to single closure (GC/SC), closure to word (C/W), and closure to following vowel (C/V2) .....	137
Table 5.5 Example of word lists and position of the target word .....	140
Table 5.6 Total number of contrast pairs in the audio clips for each participant .....	141
Table 5.7 Number of participants in each group for analysis .....	143
Table 5.8 Two principal components used in the first experiment for native Japanese listeners .....	146
Table 5.9 Two principal components used in the first experiment for L1 English-L2 Japanese listeners.....	146
Table 5.10 Two factor scores revealed by PCA for L1 English-L2 Japanese listeners.....	147
Table 5.11 Factors considered in analysis of dataset.....	148
Table 5.12 Numbers of stimuli of each place of articulation by language group.....	149
Table 5.13 Model summary for singleton-geminate discrimination by native Japanese listeners.....	152
Table 5.14 Model Summary for Singleton-Geminates discrimination by L1 English-L2 Japanese listeners .....	154
Table 5.15 Model Summary for singleton-geminate discrimination by monolingual English listeners.....	157
Table 5.16 Model summary for singleton-CVC discrimination by monolingual English listeners.....	159
Table 5.17 Model summary for singleton-CVC discrimination by L1 English-L2 Japanese listeners.....	160
Table 5.18 Model summary for singleton-CVC discrimination by native Japanese listeners	163
Table 5.19 Frequency counts of morae based on Tamaoka and Makioka (2004) .....	174

# Chapter 1

## Introduction and Literature Review

### 1.1 Introduction

This thesis focuses on the second-language acquisition of phonotactic knowledge. The aim of this thesis is to investigate whether second-language (L2) learners can learn the sublexicon phonology of a language through experience of using and/or passive exposure to the lexical stratification of a language.

Phonological knowledge is what speakers know, implicitly or explicitly, about the function and organisation of sounds in the languages they speak. One aspect of phonological knowledge is phonotactics – knowledge of the possible patterns of phoneme occurrence and phoneme sequences in a given language. Language users and learners are sensitive to distributional information, which enables them to extract regularities that occur in the language input that they are exposed to (Aslin, Saffran, & Newport, 1998; K. E. Chambers, Onishi, & Fisher, 2003, 2011; Maye, Werker, & Gerken, 2002; Onishi, Chambers, & Fisher, 2002; Saffran, Aslin, & Newport, 1996; Saffran, Newport, & Aslin, 1996).

Statistical learning refers to the process of extracting structure from one's environment (Romberg & Saffran, 2010). Native (L1) speakers are sensitive to phonotactic patterns in their language (Jusczyk, Friederici, Wessels, Svenkerud, & Jusczyk, 1993; Jusczyk, Luce, & Jan, 1994) and such phonotactic sensitivity is gradient rather than categorical (e.g., Coleman & Pierrehumbert, 1997; Frisch, Large, & Pisoni, 2000; Hay, Pierrehumbert, & Beckman, 2004; Wilson & Davidson, 2013). L1 speakers exploit their knowledge of probabilistic phonological information in the perception and production of novel words (Edwards, Beckman, & Munson, 2004; Frisch et al., 2000; Hay et al., 2004; Vitevitch & Luce, 1998; Zamuner, Gerken, & Hammond, 2004). Such knowledge is acquired by statistical learning and is based on statistical generalisation over known words (Edwards et al., 2004; Frisch, Large, Zawaydeh, & Pisoni, 2001), and it helps/guides language acquisition (e.g., Graf Estes, Edwards, & Saffran, 2011; Graf Estes, Gluck, & Grimm, 2016). However, the 'statistical learning' phonotactic literature concentrates on the learning of overall phonotactics in a language, and does not focus on cases where there might be different subsystems with different phonotactic properties within the same language.



Japanese has lexical stratification in which lexical items are classified into four strata according to their etymological status: native Japanese, Sino-Japanese (old loans from Chinese), assimilated foreign loanwords (older and more nativised loans of non-Chinese origin) and unassimilated foreign (newer and less nativised loans of non-Chinese origin) (Itô & Mester, 1995, 1999). Thus, there are etymologically defined sublexicons. Each lexicon subset shares phonological, morphological, and orthographic properties. Each stratification exhibits distinct phonological properties and certain constraints against specific segments and particular combinations of sounds (Itô & Mester, 1995). More precisely, phonotactic structures that are illicit in the native stratum are licit in the loanword (i.e., foreign) stratum. In this thesis, both the assimilated and unassimilated foreign loans are referred to as loanwords and these strata are called a loanword stratum.

Psycholinguistic studies of Japanese have shown evidence for the psychological reality of lexical stratification by adult native speakers of Japanese (Gelbart, 2005; Gelbart & Kawahara, 2007; Moreton & Amano, 1999). Lexical stratification is very salient for adult speakers of Japanese who have access to knowledge of lexical stratification in auditory perception (Gelbart & Kawahara, 2007; Moreton & Amano, 1999). Native speakers are able to intuit and detect statistical properties of sound patterns in sublexicons (Moreton & Amano, 1999). A question remains whether L2 learners have such intuitions about sublexicon phonology as generated from their entire Japanese lexicons. Recently, a sophisticated computational clustering method by Morita (2018) demonstrated that the stratal affiliation of a word in Japanese can be predicted by the segmental phonotactic probability of naturalistic data taken from a corpus. This suggests that Japanese sublexicon phonotactics are learnable by language users of Japanese regardless of whether Japanese is their L1 or not. In this thesis, I focus on a phonological/phonotactic aspect of Japanese lexical stratification and examine how L2 learners' knowledge of sublexicon phonology/phonotactics can be explained with statistical language learning. This question has important potential to advance the general understanding of phonological aspects of language acquisition.

Although there is consensus among researchers that statistical learning plays a role in language acquisition and its mechanism is used to acquire a native language, little attention has been paid to its role in L2 language acquisition. Some studies have shown that despite effects of L1 phonotactic constraints in L2 learning, advanced L2 learners are able to acquire L2 phonology and exploit their knowledge such as employing speech segmentation and detecting word boundaries (e.g., Weber & Cutler, 2006), in addition to the well-formedness of consonant clusters (e.g., Altenberg, 2005; Trapman & Kager, 2009) and prosodic structure (e.g., Preston

& Yamagata, 2004) in unsupervised contexts. However, it is not yet clear how sensitive L2 learners are to sublexicon-specific phonotactic properties, and whether they are able to track distributional patterns of observable words and generalise to novel words. Based on a large body of evidence using statistical learning mechanism paradigms on L1, stored lexical representations of individual L2 learners are reasonably assumed to reflect both the input to which they were exposed and L2 phonological grammars acquired from their lexicons.

Specifically, this study focuses on the loanword (i.e., foreign) stratum and specific phonological phenomena in Japanese loanword phonology. In a similar way to many other languages, Japanese loanword lexicons consist of words borrowed from English or other foreign languages. However, “[l]oanword phonology is closely related with and severely constrained by native phonology” (Kubozono, Itô, & Mester, 2008, p. 1). While Japanese loanword phonology contains the essential characteristics of the overall Japanese phonological system, specific rules are applied to loanwords because of the phonological and phonotactic differences between the source languages and the host language (i.e., Japanese). Therefore, loanwords exhibit specific phonotactics of the sublexicon. Since Japanese loanwords are signalled by way of a different set of orthography, *katakana* syllabary<sup>1</sup>, L2 learners might be particularly sensitive to the sublexicon phonotactics or constraints governing the language.

The current study focuses on two loanword phonological regularities. The first phonological regularity was a set of epenthetic vowels. Because Japanese phonotactics allow only a nasal or the first part of a geminate consonant in coda position (Tsuji-mura, 1996), illicit final consonants and consonant clusters in the source language are modified by inserting an epenthetic vowel in borrowed words. One of three different vowels, /i, o, u/ is selected to be an epenthetic vowel, depending upon the preceding consonant, which reflects co-occurrence phonotactic restrictions on CV sequences of CV syllables in native-Japanese (Hirayama, 2003; Irwin, 2011; Kubozono, 2015). The default epenthetic vowel is /u/. Other two vowels /i/ and /o/ are epenthesised only in certain contexts. For example, ‘pink’ is borrowed as *pinku* whereas ‘tent’ is borrowed as *tento*, since the [tu] sequence is illicit in native-Japanese. Epenthesis obligatorily occurs, since closed syllables (i.e., with codas) are disallowed unless the coda consonant is nasal. In addition, the quality of the epenthetic vowels is predictable as the choice of epenthetic vowel is systematic and categorical. I assume that categorical rules are more easily learned than gradient phonotactics (e.g., Shea & Curtin, 2011). Therefore, epenthetic vowels

---

<sup>1</sup> *Katakana* is phonographic and a syllabary; each letter corresponds to one mora. *Katakana* is basically used for loanwords and mimetic words. See §2.2.2 for details of orthographic system in Japanese.

must be expected be learnable from language exposure when learners encounter Japanese loanwords whose source words are their first language.

The second phonological regularity of loanwords evaluated is consonant gemination. The process of epenthesis is often accompanied by the process of obstruent consonant lengthening (i.e., gemination), when coda consonants follow a lax vowel (Kubozono, 2015). That is, final stops following lax vowels in English CVC words are borrowed as geminates including voiced stops (Itô, Kubozono, & Mester, 2017; Shirai, 1999; Takagi & Mann, 1994). For example, ‘pet’ is borrowed as *petto* and ‘head’ is borrowed as *heddo* rather than *hedo*. While voiced geminates are permitted in loanwords, they are phonotactically constrained in native phonology. That is, geminate obstruents must be voiceless in all Japanese sublexicons, except that of loanwords. However, in nativised loanwords, voiced stops do not undergo gemination as frequently as voiceless stops. In addition, voiced geminates in loanwords exhibit place asymmetry (Hirayama, 2005). While [dd] quite frequently appears in loanwords, [bb] is rare and [gg] falls between these two (i.e., [dd] > [gg] > [bb]) (e.g., Shirai, 1999, Hirayama, 2005, Amano & Kondo, 2000 cited in Kawahara, 2005). (Hereafter, the Japanese long vowels and consonants will be transcribed such as [aa], [ii], [kk], [gg].) Thus, the frequency of voiced geminates in loanwords is strongly related to the place of articulation. Therefore, statistical properties of sound patterns in the sublexicon can be hypothesised to have different degrees of well-formedness in Japanese loanwords. Hay et al. (2004) show that frequencies of clusters constrained by a homorganic rule in the lexicon received gradient speech perception and well-formedness judgments. Well-formedness is related to type frequency (Hay et al., 2004). The patterns in gemination in loanwords enable us to examine whether learners of the language have gradient well-formedness judgments of voiced geminates in Japanese loanwords.

In summary, epenthetic vowels are a phenomenon only existing within the loanword lexicon<sup>2</sup>, and there are categorical constraints dictating which vowels should be used. For geminates within other sublexicons (native and Sino-Japanese words), there is a categorical constraint preventing them. However, within our target sublexicon (i.e., loanwords), there are stochastic patterns, such that the likelihood of different voiced geminates varies. Moreover, voiced geminates are less likely than voiceless geminates even in loanwords. These regularities are not taught in standard curricula for both L1 and L2 Japanese<sup>3</sup>.

---

<sup>2</sup> This is debatable. Some scholars argue that epenthesis occurs even native Japanese (e.g., Poser, 1984) and Sino-Japanese (e.g., Itô & Mester, 1996, 2015).

<sup>3</sup> I have discussed these regularities with a faculty member of the Department of Japanese in the University of Canterbury whether students are taught in standard curricula before proposing this study.

It is a question whether by mere exposure to Japanese, language learners can learn such phonological phenomenon and stochastic patterns that only exist within the loanword lexicon. In addition to the above regularities in loanwords, the phonotactic scores of overall Japanese and that of loanwords in experimental stimuli were calculated based a dictionary corpus created from the data of the Corpus of Spontaneous Japanese (National Institute for Japanese Language and Linguistics, 2017). As part of the analysis, this thesis investigated whether loanword phonotactics and overall phonotactics make separate contributions to perceived well-formedness. In order to address these questions, a confidence-rating task (which is a type of a well-formedness judgment task) was conducted on the adaptation of English final consonants in monosyllabic words by three language groups: native speakers of Japanese, English-speaking learners of Japanese and English-speaking non-learners of Japanese. The ratings would reflect their knowledge of the loanword phonology in Japanese produced by statistical learning and stochastic patterns of borrowings. Because categorical rules and stochastic patterns (i.e., gradient rules) in loanwords are considered, this study helps us to better understand what knowledge related to Japanese loanword phonology L2 learners have readily learnt. This research also considers the degree of language exposure experienced by the individual. If learning the sublexicon phonology is possible, then acquisition would be different depending on the degree to which a learner is exposed to Japanese. This is based on an assumption that participants who have more exposure to Japanese have a reasonable level of phonological knowledge within their accumulated Japanese lexicon. In this study, the participants filled in a questionnaire which assessed their degree of exposure to Japanese. The research questions and hypotheses are presented in more detail in §2.4.

On the basis of the results of the well-formedness task, the second part of this thesis explores auditory discrimination of non-native sound contrasts. The question raised was whether there is perceptual confusion between single and geminate consonants by L2 learners. In order to help determine whether L2 learners tap into phonotactic knowledge during the well-formedness judgements and whether they perceive non-native segmental contrasts in spoken words, an auditory memory decision task was conducted.

The rest of this thesis is organised as follows: in Chapter 1, an overview of the statistical learning phonotactic literature into native speech and the effect of vocabulary size is discussed (§1.2). In §1.3, I review some of the rule-learning literature and in §1.4, the L2 acquisition of phonotactics is also discussed in relation to the effects of first language phonotactics in L2 learning, and the experimental literature is reviewed. A knowledge gap in the field of study is discussed.

Chapter 2 introduces lexical stratification of Japanese lexicons in §2.1 and Japanese phonology in §2.2. The existing empirical evidence for the lexical stratification by native speakers of Japanese is also discussed. Then, loanword phonology in Japanese is introduced in §2.3. Specifically, the strategies used to select epenthetic vowels in adapted foreign words and the process of consonant gemination are examined. This is followed by a corpus work which presents stochastic patterns of the gemination of loanwords in Japanese, collected based upon a large-scale corpus, which reflects the actual language use of loanwords in Japanese. Then, the research questions and hypotheses are presented in §2.4.

Chapter 3 provides the materials and methodology used in the pilot studies and a brief overview of the outcomes from pilot studies and adaptations to the final research project. The main aim of the pilot studies was to validate the research method and to identify issues before the actual study.

Chapter 4 presents a fully-crossed auditorily perceptual experiment, a well-formedness judgment task which shows the adaptation of English final consonants in monosyllabic words. The experiment explored whether listeners have the loanword phonology and sublexicon phonotactics by comparing the three groups: native speakers of Japanese, English-speaking learners and non-learners of Japanese. Japanese listeners demonstrated gradient well-formedness judgements on voiced geminates for adaptation of English words. Although L2 learners' performance has not achieved the level of native speakers in case of the gemination, they have some knowledge of the adaptations needed for loanwords in Japanese and show sensitivity to the quality of epenthetic vowels. Moreover, participants' well-formedness ratings are more related to the probability of the loanword phonotactics than that of the Japanese overall phonotactics.

Chapter 5 presents another cross-linguistic experiment, an auditory memory decision task builds on the results of the first experiment to explore the role of perceptual discrimination of non-native consonant length contrasts, by comparing native speakers of Japanese, English-speaking learners and non-learners of Japanese. The first object is to investigate whether L2 learners of Japanese tap into phonotactic knowledge during the well-formedness judgements. Specifically, this second experiment is designed to examine the degree to which non-native speakers can perceive contrasts that do not occur in their native language – singleton/geminate contrasts for English speakers, and CV/CVC contrasts for Japanese speakers. It also investigates the degree to which success in this task is mediated by phonetic salience of the particular contrast, and by the individual's previous language experience. To anticipate the results, non-native listeners' performance has not achieved the level of native speakers on non-

native sound contrast. In particular, L2 learners of Japanese had different degrees of perceptual discrimination depending on phonetic salience in stimuli and across learners. However, it did not indicate that L2 learners of Japanese were not able to discriminate between singleton and geminates.

In Chapter 6, experimental findings are summarised, and the implication of the findings and constraints on statistical learning are discussed along with some limitations and future directions. Then, a general conclusion is presented.

## **1.2 Statistical Learning and Acquisition of Phonotactics**

### **1.2.1 The effects of statistical knowledge on word recognition, production, and perception in L1s**

A considerable amount of literature has explored the influence of statistical learning in different areas of linguistics (e.g., K. E. Chambers et al., 2003; Gómez & Lakusta, 2004; Maye et al., 2002; Onnis, Waterfall, & Edelman, 2008; Pacton, Perruchet, Fayol, & Cleeremans, 2001; Saffran & Wilson, 2003). In terms of phonological acquisition, there is broad consensus among scholars that L1 speakers of a given language, both infants and adults, are sensitive to their language's phonotactics, phonological rules and transitional probability of phoneme combinations, and such knowledge affects language processing (e.g., Coleman & Pierrehumbert, 1997; Jusczyk et al., 1994; McQueen, 1998; Otake, Yoneyama, Cutler, & Van Der Lugt, 1996; Vitevitch, Luce, Charles-Luce, & Kemmerer, 1997; Yip, 2015).

For instance, a word-spotting experiment by McQueen (1998) shows that knowledge of the phonotactic probability of onsets and codas facilitated Dutch adult listeners to detect embedded Dutch words in auditory bisyllabic nonsense stimuli. In Dutch phonotactics, phoneme sequences [lv] and [mr] cannot occur within a syllable. Dutch listeners exploited the knowledge of the sequencing constraints which cues syllable boundaries as well as possible word boundaries for lexical segmentation. Thus, *pil* 'pill' was detected faster and more accurately in stimulus words with phonotactic alignment like [pɪl.vrem], than in words without alignment like [pɪlv.rem]. Similarly, *rok* "skirt" was detected faster and more accurately in [fɪm.rɔk] than in [fɪ.drɔk]. The effects of sequencing constraints resulted in successful parsing speech segmentation into real words.

Such sensitivity to phonotactics in the L1 is sharpened in early development. Many studies show that infants exhibit early sensitivity to possible sound patterns in their native languages before they begin producing words (e.g., Jusczyk et al., 1993; Jusczyk et al., 1994; Mattys & Jusczyk, 2001; Mattys, Jusczyk, Luce, & Morgan, 1999). Infants between six to eight-

months-old are able to discriminate their native phonemes from non-native sounds (Trehub, 1976; Werker & Tees, 1984), and during the first year of life, infants are able to detect not only individual segments but also sequential patterns present in the ambient language (Friederici & Wessels, 1993; Jusczyk et al., 1993).

While language users are sensitive to licit or the probability of phoneme sequences in their language and exploit their phonotactic knowledge for speech processing, they also show sensitivity to illicit or unattested sequences. In the traditional view of generative phonology, phonemes and phoneme sequences are assumed to be categorical between phonologically legal and illegal structures. Therefore, all attested phonemes and phoneme sequences are treated as equally legal, whereas unattested phonemes and phoneme sequences are considered equally illegal, as if they are uniform (Ernestus, 2011). However, L1 speakers' well-formedness judgements are gradient depending on the phoneme combinations, even if such combinations are unattested (Albright, 2009; Coleman & Pierrehumbert, 1997; Hay et al., 2004). Empirical studies on well-formedness judgements show that speakers' phonotactic knowledge reflects frequency of the target segments in the lexicon, and phonotactic constraints are gradient rather than categorical (Bailey & Hahn, 2001; Coleman & Pierrehumbert, 1997; Frisch et al., 2000; Hay et al., 2004; Vitevitch et al., 1997; Wilson & Davidson, 2013). Gradient acceptability or judgements are observed not only within rating experiments but also within categorical binary choice experiments (e.g., Frisch et al., 2000; Kawahara, 2010).

*Probabilistic phonotactics* refers to the relative frequencies of the occurrence of sounds and sequences of sounds in syllables and lexicons (Vitevitch & Luce, 1999). Frequently appearing phonotactic patterns influence production, recognition, and processing of novel spoken stimuli (Edwards et al., 2004; Hay et al., 2004; Jusczyk et al., 1994; Vitevitch & Luce, 1998, 2005; Zamuner et al., 2004). Sensitivity to probabilistic phonotactics in the native language is developed in the speaker's first year. Nine-month-old infants preferred to listen to words with higher probabilities of the phonotactic sequential patterns in their native language compared to those with lower probabilities of sound patterns (e.g., Jusczyk et al., 1994). In addition to infants, children and adults also show this similar effect of phonotactic probabilities. Nonword stimuli that contain highly frequent sound sequences were produced faster and more accurately (Edwards et al., 2004; Vitevitch & Luce, 1998, 2005; Zamuner et al., 2004), were judged more word-like (e.g., Coleman & Pierrehumbert, 1997; Frisch et al., 2000; Vitevitch et al., 1997), and were more easily recognised (Frisch et al., 2000) than nonwords that contained less frequent phonotactic patterns. In addition, such effects are observed even transcribing

nasal-obstruent sequences in nonwords. The low frequency sequence /np/ embedded in nonwords was transcribed as a more frequent sequence /mp/ (Hay et al., 2004)

To summarise, language speakers are sensitive to phonotactic patterns in their own language, which influences the strategies speakers use to segment words from the speech stream. Speakers' well-formedness judgments on nonwords contained phonotactically licit/illicit sequences that are gradient rather than categorical. Moreover, there is relation between well-formedness and the statistics of the lexicon. Words with higher probabilities of the phonotactic sequential patterns are processed more accurately and faster than words with low probability sequential patterns. Empirical studies indicate that native speakers are able to exploit their knowledge of probabilistic phonotactic information in perception and production of novel words through statistical language learning. Such ability promotes language acquisition. However, while the 'statistical learning' phonotactic literature concentrates on the learning of overall phonotactics in a language, cases where languages have different subsystems with different phonotactic properties within the same language have been paid less attention. This issue will be discussed in more detail in relation to the Japanese language, in §2.2.3, using empirical studies.

## **1.2.2 The effects of vocabulary size on statistical learning**

In addition to the relative frequency of the sounds having an effect on extracting phonotactic information from language inputs, the impact of speakers' vocabulary size on phonotactic acquisition is also a factor, related to learning probabilistic sequences (Edwards et al., 2004; Frisch & Brea-Spahn, 2010; Frisch et al., 2001; Graf Estes et al., 2011; Graf Estes et al., 2016; Pierrehumbert, 2001; Storkel, 2001; Storkel, Armbrüster, & Hogan, 2006; Storkel & Hoover, 2011).

Storkel (2001) examined the influence of phonological probabilities on novel word learning by children aged between three to six years, by using two sets of nonwords (CVC) which varied in their phonological probabilities: common vs. rare. Children learned nouns with common sound sequences faster and more accurately than words with rare sequences across different measures of learning. Furthermore, these tendencies are more likely to increase as children' receptive vocabulary increased. Note that in general, language users' receptive vocabulary is generally related to their comprehension and listening skills, whereas measures of productive (or expressive) vocabulary size helps determine how language users are able to speak or write (Webb, 2008). Based on these findings Storkel suggests that successful language



learners seem to have a larger receptive vocabulary which facilitates the gathering of phonotactic information, and this phonotactic knowledge allows children to acquire more new words.<sup>4</sup>

A later study by Graf Estes et al. (2011) reinforced the findings of Storkel (2001) about the correlation between the knowledge of phonotactics and the size of receptive vocabularies. Moreover, in this study, 18-month-old infants exhibited phonotactic constraints on novel word learning. That is, while infants with larger receptive vocabularies (above the median receptive vocabulary size in participated infants: 303 words) tend to successfully learn phonotactically licit stimuli, they struggled to learn more illicit stimuli than infants with smaller vocabularies (below 303 words). The group with smaller vocabularies did not exhibit substantial differences between learning legal and illegal stimuli. This study revealed that novel word learning is influenced by knowledge of native phonotactics that facilitates learning phonotactically licit vocabulary, but constrains the learning of phonotactically illicit words.

The effects of vocabulary size on adult perception of wordlikeness is also found in Frisch et al. (2001). Sequences with varying low frequencies in the lexicon are accepted in different degrees of wordlikeness judgments by adults with greater lexical knowledge. This is because less frequent/probable items are more likely to occur in a larger lexicon than in a small lexicon. Adults with less lexical knowledge treated all low probability sequences the same way.

Thus, the vocabulary size takes into account the individual differences in the well-formedness judgments of the participants. Taken together, with respect to L1, the effect of individual speakers' vocabulary size on word learning seems to be robust as reported above. When considering sublexicon phonology in a language, we need to acknowledge the vocabulary size of sublexicons as well as a speaker's overall vocabulary size. This issue will be discussed with data from a corpus in §2.3.3.

### **1.3 Rule Learning in Artificial Languages**

In regard to the studies of statistical learning, Aslin (2017, p. 6) states “the original idea proposed by Saffran et al. was that learners rely, at least in part, on computing transitional probabilities between adjacent syllables. But a variety of other models have been proposed over the past 15 years .... [a] variety of models exhibit the gradient property of generalization from

---

<sup>4</sup> Note that in contradiction to this study, the relation between the effect of phonotactic probabilities and receptive vocabulary size was not found in verb studies in Storkel (2003). This discrepancy might be attributed to the small effect of phonotactic probability on verbs and to different word learning paradigm between nouns and verbs.

specific exemplars (statistical learning) to abstract principles (rule learning)”. Language learners are not only sensitive to distributional information in the language input that they are exposed to, but they are also able to use the information to generalise the learned or experienced patterns to novel exemplars in the artificial language studies (e.g., Gerken, 2006; Gómez & Gerken, 1999; Marcus, Vijayan, Rao, & Vishton, 1999; Maye et al., 2002; Mintz, 2002; Reeder, Newport, & Aslin, 2013, 2017). This ability or abstract process is referred to as rule learning (Aslin, 2017, p. 5).

Marcus et al. (1999) examined the acquisition of rule learning which cannot be accounted for by statistical information such as transitional probabilities. They investigated whether 7-month-old infants are able to generalise the repetition rule to novel words by using 16 sets of nonsense 3-syllable-strings (e.g., AAB, ABB or ABA such as *leledi*, *ledidi*, *ledile*). Both the A and B elements were selected from four different types of syllable, respectively. Thus, each set differed one from another. During a 2-min familiarisation phase, infants were randomly assigned to two conditions (e.g., either AAB condition or ABB condition) and they listened to 16 strings. In the test phase, for half of the test trials, infants were exposed to completely novel syllables that were consistent with familiarised strings, whereas for the other half, novel syllables inconsistent with familiarised strings were presented the infants. For example, infants in the ABB conditions were trained for the ABB grammar. Therefore, for these infants, the AAB grammar is not consistent with their familiarised grammar (i.e., ABB). The results show that the infants looked longer at the flashing light for the inconsistent grammar than for the consistent grammar, indicating that the infants discerned the unfamiliar grammar from the familiar grammar. From the findings Marcus et al. argue that the learning reflects the infants’ ability to extract learned structural patterns or abstract rules. This is because infants cannot rely on statistical cues as unfamiliar syllable strings do not provide transitional probabilities to them.

However, Aslin and Newport (2012, 2014) and Aslin (2017) claim that statistical learning and rule learning are based on the same single domain-general mechanism, with a gradient of generalisation rather than separable mechanisms. Some studies of artificial grammar learning support this view. For example, Gerken (2006) considered findings from artificial language-learning literature that infants are able to discern given structures by providing sufficient distributional information in the input to utilise for generalisation. In the stimuli of Marcus et al. (1999), there were two types of AAB strings, in that (1) the B element varies among stimuli (e.g., *leledi*, *wiwije*, *jijili*, *dedewe*) and (2) the B element is always the same syllable (e.g., *leledi*, *wiwidi*, *jijidi*, *dededi*). Thus, two different AAB rules can be generalised

to either duplicate the first syllable or end with a specific syllable (i.e., AAdi). In addition, if infants can generalise the ABB rule in condition (1), they need to utilise only four stimuli instead of the entire 16 stimuli. Therefore, Gerken investigated whether 9-month-old infants generate a rule based on given structures: either a more abstract rule (i.e., AAB rule) or a specific rule (i.e., AAdi rule), using 4 of the 16 strings. During the 2-min familiarisation phase, infants were randomly assigned to either four AAB or ABA strings. Half of the infants were familiarised with the strings of the more abstract condition, and the other half were exposed to the strings of the specific rule condition. In the test phase, infants listened to four strings (2 AAB and 2 ABA). As a result, the former group were able to generalise the AAB/ABA rule to novel test stimuli. However, the latter group were not able to generalise to novel stimuli because the infants learned the subset specific rule AAdi/AdiA rather than the more abstract rule in which B can be variable. A subsequent experiment confirmed that the infants in the latter group generalise the structure associated with the position of the syllable *di*. Importantly, this study shows that the generalisation made by infants reflects the likelihood of structural patterns of data provided.

In a follow-up study, Gerken (2010) found that adding three counter examples into the specific condition enabled infants to shift from the specific rule to the more abstract rule, suggesting that infants consider multiple possible models appropriate for the data provided. Thus, the findings are in line with the view that statistical learning and rule learning share the single general mechanism which enables learners to extract the statistics of given inputs, according to common patterns in the input, whereby speakers use this information to generalise during unsupervised learning.

Like infant learners, adults are also capable of utilizing distributional cues to discern abstract categories without phonological or semantics cues (e.g., Mintz, 2002; Mintz, Wang, & Li, 2014; Reeder et al., 2013, 2017). For example, in a series of experiments, Reeder et al. (2013) investigated whether adults generalise learned artificial grammatical categories for new sentences solely based on distribution contexts for words. The grammar used in the experiments consisted of three-to-five word sentences (Q)AXB(R), in which each letter refers to a grammatical category of nonsense words: 3 A-words, 3 X-words, 3 B-words, 2 Q and 2 R-words. Adult learners were exposed to the artificial grammar under different conditions in which three distributional variables are manipulated: density, overlap and frequency. Density refers to the number of different contexts in which each word in the input set occurs, and overlap denotes how much contextual information overlaps across the target X-words, whereas frequency refers to the amount of exposure to these cues. In the test phase, learners listened to

grammatical familiar, grammatical novel, and ungrammatical novel sentences (e.g., AXA, BXB). Reeder et al. found that adult learners are sensitive to the distributional information embedded in the inputs according to the different conditions to which learners were exposed. When contexts overlap across different target words, learners utilise this information and generalise the target category *X* to novel sentences. That is, when stimuli contain a complete overlap of contexts across words, novel grammatical sentences were more highly rated than when stimuli contained incomplete overlap in the sets of contexts. Moreover, when learners were exposed to the stimulus sets containing incomplete overlap used in the previous experiment thrice rather than once, they are less likely to generalise to new sentences. Thus, category learning shows graded effects, as adult learners determine whether to generalise according to distributional information provided, especially by the degree of the overlap of contexts across words. Subsequent study (Reeder et al., 2017) showed that adult learners even acquire subcategories.

In sum, in terms of artificial grammar learning, language users are capable of detecting distributional patterns in systematic ways without other additional information. Rules are defined broadly or narrowly according to the distributional information given. This is because the generalisation by infants and adults mirrors the learned or exposed patterns in their inputs. Word co-occurrence statistics are exploited by learners, suggesting statistical and rule learning mechanisms are accounted for by the same mechanism. This thesis treats statistical and rule learning on the basis of this view by Aslin and Newport (2012, 2014).

The important thing is that statistical and rule learning mechanisms might play a role in L2 acquisition. Natural languages are more complicated especially when a language has different subsystems with different phonotactic properties within the same language. As a reminder, this dissertation explores L2 acquisition of Japanese sublexicon phonology. Although Japanese loanword phonology is explained in more detail in Chapter 2, epenthetic vowels are a phenomenon existing only within the loanword lexicon, and there are categorical constraints dictating which vowels should be used. For geminates, within other sublexicons (native and Sino-Japanese words), there is a categorical constraint preventing them. However, within our target sublexicon (i.e., loanwords), there are stochastic patterns, such that the likelihood of different voiced geminates varies. Moreover, voiced geminates are less likely than voiceless geminates in this context. I assume that both forming general rules and detecting stochastic patterns are used via the same mechanism as discussed above. The question of interest is whether language users can extract the patterns of epenthetic vowels from instances of distributional contexts of loanwords in the natural language, generalizing patterns to novel

instances without supervision. The other question is whether learners are sensitive to the fine-grained patterns of voiced geminates that only occur in sublexicons such as loanwords. I will return these questions in more detail in §2.4 after discussing the Japanese language.

#### **1.4 Acquisition of L2 Phonotactic Knowledge**

As seen in the previous section, a large and growing body of literature has investigated statistical learning in L1 phonotactics, reporting its effect on word recognition, production, and perception. Speakers are sensitive to probabilistic phoneme sequences and have fine-grained phonotactic knowledge of their languages. However, most studies of L2 acquisition related to phonetics and phonology have focused on the pronunciation ‘accuracy’ of individual phonemes and sequential phonemes that are not attested in the phonological inventory of the learners, or have looked at the ‘interference’ of learners’ native phonological system, including phonotactic knowledge, on pronunciation and perception in non-native sound sequences (e.g., Best, 1994, 1995; Best, McRoberts, & Goodell, 2001; Flege, 1991, 1995; Kuhl, 1993; Kuhl et al., 2008). Studies on the perception of non-native sound consistently show that non-native sound structure is perceptually assimilated into the licit structure in a speaker’s native language (e.g., Dehaene-Lambertz, Dupoux, & Gut, 2000; Dupoux, Kakehi, Hirose, Pallier, & Mehler, 1999; Hallé, Segui, Frauenfelder, & Meunier, 1998; Kabak & Idsardi, 2007; Massaro & Cohen, 1983), suggesting that phonotactic knowledge of listeners’ L1 regulates how non-native sounds are perceived. Such knowledge seems to be interfering with the development of perception and production in L2. In addition, these studies indicate that L2 learners have difficulty in acquiring phonological knowledge in the target languages.

There has been relatively little attention paid to the relations between the psychological realities of L2 phonotactic knowledge (i.e., metalinguistic judgement), perception and production performance. Although much uncertainty exists about the development of phonological knowledge including acquisition of L2 phonotactics in comparison to that of L1, some empirical studies examined the subconscious grammatical knowledge of L2 learners. The studies reviewed below indicate that L2 listeners access and make use of the language-specific probabilities of an L2 in different tasks such as word well-formedness judgements and speech segmentation. These studies show that L2 learners derive implicit L2 phonotactic knowledge about consonant clusters that are not attested in the native language.

Altenberg (2005) investigated the acquisition of phonotactics in L2 learning by comparing 30 Spanish-speaking L2 learners of English (university students) and 10 native

speakers of English. In Spanish, /sC/ clusters are not allowed. Therefore, native speakers of Spanish tend to epenthesise a vowel before English word-initial /s-/. In the study, it is assumed that the metalinguistic knowledge of learners and their implemented knowledge in performance might be different. Therefore, three distinctive methods were used: well-formedness judgments, perception, and production of English /sC/ onsets. In the well-formedness rating task, three types of nonwords were presented orthographically (as considered different phonetic realisation of the same phoneme) to participants as new words of English and Spanish: type ES contained initial clusters that are grammatical in both English and Spanish (e.g., /fl, dr, kr, bl/); type E\*S contained initial clusters that are grammatical in English but not in Spanish (e.g., /sp, sm, sn, sl/), and finally type \*E\*S contained initial clusters that are ungrammatical in both English nor Spanish (e.g., /sr, zn, dl, fn/). On the one hand, native and non-native speakers assessed the acceptability of nonwords as English new words on a scale 1 (completely acceptable) to 5 (completely unacceptable), as instructed to do so in English. On the other hand, non-native speakers also assessed the acceptability of nonwords as Spanish new words, following instructions in Spanish. In the English version of the task, the results revealed no significant difference between native and non-native English speakers for type ES and type E\*S. Importantly, non-native speakers judged ES and E\*S nonwords as significantly more acceptable than \*E\*S nonwords. In addition, there was no effect from the level of L2 proficiency according to the class they were placed in by the university. This suggests that the L2 learners acquire knowledge of English phonotactics of onset clusters regardless of their L2 proficiency. However, between the two versions, native Spanish participants' ratings for E\*S and \*E\*S nonwords on the Spanish version of the task were significantly higher than on the English version, indicating that participants used different phonological knowledge to judge the acceptability for each language.

In a perception task, participants were asked to listen to nonwords, and to then write down the consonant cluster they heard. Since no relation was found between accuracy and type of words (i.e., ES and E\*S), effect of transfer was not detected, suggesting learners used phonotactic knowledge of the L2. Lastly, for the production task, there was significant effect related to the types of words, which meant that native speakers of Spanish made more errors on E\*S than ES words. Altenberg interprets her results as indicating that there was no evidence of the effects of native language phonotactics on the metalinguistic task and perception of onset clusters. The L2 learners are able to acquire phonological knowledge of L2. This study also indicates disassociation of metalinguistic knowledge and oral production.

Weber and Cutler (2006) investigated L1 phonotactic constraints in L2 listening, by comparing response patterns of native speakers of German with high proficiency in English as L2, and native speakers of English with no knowledge of German. The L2 learners who participated in this study were 48 students of English translation and interpretation at a university in Germany, and they had experience in learning English, with 15 years on average, beginning at a mean age of 11. A word-spotting task was used to examine whether participants were able to detect embedded English words in spoken nonword stimuli (e.g., *thrarshlecture*, *glarshwish*). Phonotactic sequencing constraints differ between English and German. For instance, /ʃl/ and /ʃw/ are possible onset clusters in German words, but not in English, whereas /sl/ and /sw/ are possible onset clusters in English, but not in German. In order to compare detection of the same word in different contexts, four distinctive preceding contexts were created to provide clear boundary constraints: both English and German boundary (e.g., *moinlecture*), an English-only boundary (e.g., *thrarshlecture*), a German-only boundary (e.g., *moycelecture*), and neither language (e.g., *gorklecture*). The results showed some evidence that L2 English learners were able to exploit the phonotactic probabilities of English in order to detect word boundaries. That is, the English-only boundary constraint facilitates word identification for German listeners almost as strongly as it does for native English listeners. German listeners acquire knowledge of English illicit clusters. However, for the German-only boundary (e.g., *moycelecture*), German listeners also exploited the phonotactics to spot the embedded word, suggesting L1 phonotactic constraints influence L2 listening. Even so, this study shows that such constraints affect speech segmentation, and advanced L2 learners are capable of learning phonotactic probabilities of the L2. Their findings raise a question as to whether L2 learners who started learning the target language after adolescence are able to learn L2 phonological constraints including phonotactics. From the information, it can be speculated that the speakers have large vocabularies of English.

Similarly, Trapman and Kager (2009) found that advanced L2 learners can acquire L2 phonotactic knowledge and such knowledge is subject to development. They examined L2 acquisition of Dutch consonant clusters in relation to a subset and superset of Dutch phonotactics. Consonants cluster used in their experiment are (1) attested in Russian only, (2) attested in Russian and Dutch but not in Spanish, and (3) attested in all three languages. Thus, Dutch is both the subset and the superset grammar. Russian learners of Dutch as subset learners, Spanish learners of Dutch as superset learners, and native speakers of Dutch all performed word-likeness judgements for nonwords containing clusters in onsets and codas, on a seven-point scale. Each learner group consisted of advanced and beginning learners. Trapman and

Kager found that Russian learners of Dutch assigned significantly higher ratings to Dutch-attested clusters in nonword stimuli compared to Dutch-unattested clusters, even though Dutch clusters are a subset of those in Russian. Taking the results, Trapman and Kager were concerned that results were influenced by the Russian lexical statistics on the Russian learners' responses. However, while the correlation between the Russian bi-phone probabilities of the stimulus words and average word-likeness judgments of the stimulus words is not significant, the Dutch bi-phone probabilities correlate significantly with the ratings of Russian learners. Hence, judgments of Russian learners are more likely to be derived from their phonotactic knowledge of Dutch rather than statistical based lexical similarities between Russian and Dutch. For Spanish learners, beginning learners did not distinguish the unattested onset clusters from the attested ones, but they could distinguish the differences for the coda clusters as effectively as advanced learners. For native speakers of Dutch, phonotactic judgements were gradient rather than based upon categorical knowledge within the broad distinction of legal and illegal consonant clusters. Russian and Spanish advanced learners also made similar judgements to native Dutch speakers, suggesting they have native-like phonotactic knowledge. In sum, while Russian learners were aware that their attested clusters are illicit in Dutch, Spanish advanced learners also know their unattested clusters are licit in Dutch. Thus, superset and subset learners are very likely to acquire phonotactics of the target language.

In sum, these studies reveal advanced L2 learners' phonotactic sensitivity towards their L2, indicating that they know some knowledge of L2 phonotactic constraints. Advanced learners are more likely to acquire L2 probabilistic phonotactics, which suggests that the more learners are exposed to the target language, the more successfully they acquire L2 phonotactic knowledge. However, languages in the studies commonly use an alphabet writing system and thus, learners might gain advantage from orthographic inputs for learning licit/illicit clusters in L2 phonotactics. Especially, nonwords stimuli in Altenberg's well-formedness rating task were presented orthographically to participants. However, the study by Lentz and Kager (2015) also showed that acquisition of L2 probabilistic phonotactic in Dutch is possible by Japanese learners of Dutch. Knowledge of phonotactic constraints of L1 does not always hinder the ability to acquire the phonotactic probabilities of L2. Thus, L2 learners are indeed capable of detecting L1 illicit clusters as being licit in the target language

In terms of a sublexicon phonology, Preston and Yamagata (2004) investigated what kind of strategies native speakers of Japanese and English L2 learners of Japanese use during orthographical adaptation of English loanwords in Japanese. The L2 learners group consisted of 119 native speakers of English studying Japanese at universities in the United States. In



nativised English loanwords of Japanese, when English words of CVC syllable structure contain a single voiceless stop in word-final position preceded by a lax vowel, those singletons are often adapted as geminates in Japanese (e.g., pet [pet]→ [petto]) (Katayama, 1998; Koo & Homma, 1989; Shirai, 1999). In order to find out the relationship between the segmental structure of the source words and their phonological representation, participants were given the written stimuli such as *tap*, *cot*, and were asked to write the English words in *katakana* syllabary. Findings indicate that while L2 learners geminated less than native speakers, the occurrence of gemination increased in relationship to their level at university. However, how often entire words or coda consonants were correctly modified are unknown. Most interestingly, findings showed that L2 learners used long vowels as an adaptation strategy instead of germination, regardless of their grades, which indicates they are also sensitive to the number of morae in the realisation of the loanwords (i.e., *tap*: gemination [tap.pu], or vowel length [taa.pu]). In addition, both L2 learners and native speakers of Japanese tend to avoid trimoraic syllables like \*[taap.pu], which is disallowed in Japanese.

Preston and Yamagata (2004) speculated that the acquisition of CV constraints in Japanese precedes the acquisition of specific phonological patterns. When considering the overall Japanese vocabulary, the majority of syllables in Japanese are a single vowel (e.g., /a/, /e/, /o/) or CV (e.g., /ka/, /te/), whereas independent morae that differentiate between a mora and a syllable such as the nasal /N/, the first part of geminate (or voiceless obstruent) and the long vowel do not frequently occur. Especially, the ratio of geminates are smaller than that of long vowels. Therefore, learners who transcribed CVC words to CVCCV show a kind of sensitivity for the observed patterns in loanwords. In addition, the size of vocabulary and amount of speaker's exposure to the target language seems to relate to statistical learning of the observed phonological pattern in the target language; as the grade goes up, the correct modification of English words increased. This is consistent with other L2 studies discussed above.

However, it is not yet clear how sensitive L2 learners are to phonemic sequences that appear only in a sublexicon and its phonology originating from Japanese. All studies discussed above show that advanced learners are able to acquire L2 phonotactics, suggesting that the size of vocabulary plays a role. In general, as the grade goes up, learners are more exposed to the target language, building up their vocabulary. However, when we consider sublexicons, the size of vocabulary is apparently smaller than the entire vocabulary of the target language. Are learners able to acquire a sublexicon-specific phonology? Then if it is possible, acquisition of a sublexicon phonology would differ depending on the degree to which a learner is exposed to

the target language. This is based on a general assumption that learners who have more exposure to the target language have a reasonable level of phonological knowledge with their accumulated sublexicons. The psychological reality of the sublexicon phonotactics in Japanese was empirically examined by L1 Japanese speakers, who showed their sensitivity and intuition towards specific phoneme or co-occurrence of phonemes (Gelbart & Kawahara, 2007; Moreton & Amano, 1999). One remaining question is whether L2 learners have intuitions about sublexicon phonology generated from their entire Japanese lexicons. I will address more detailed research questions connected to this in §2.4. In order to address the research questions, we need to first discuss Japanese language in relation to its lexical stratification, which is followed by reviews of Gelbart and Kawahara (2007) and Moreton and Amano (1999).

# Chapter 2

## Lexical Phonology of Japanese

### 2.1 Lexical Stratification of the Japanese Language

Because of its historical development, Japanese has lexical stratification in which lexical items are classified into four strata according to their etymological status: native Japanese, Sino-Japanese (old loans from Chinese), assimilated foreign (older and more nativised loans of non-Chinese origin) and unassimilated foreign (newer and less nativised loans of non-Chinese origin) (Itô & Mester, 1995, 1999, 2001). This stratification corresponds to a different historical source of lexical items and plays a significant role on the phonological constraints that apply only to each stratum. Itô and Mester (1995, 1999, 2001, 2008) propose the Core-Periphery model of the lexicon, that represents the subsets of synchronic lexical items based on markedness constraints of the language. Itô and Mester (1995) state “several phonological constraints are stratum-specific and hold only for a particular morpheme class” (p. 819). Figure 2.1 illustrates lexical stratification based on Itô and Mester (1999), in which a domain of entire lexicons includes four smaller sublexicon domains: native, Sino-Japanese (established loans), assimilated foreign and unassimilated foreign. Importantly, native lexicon is considered to be a subset of other lexicons.

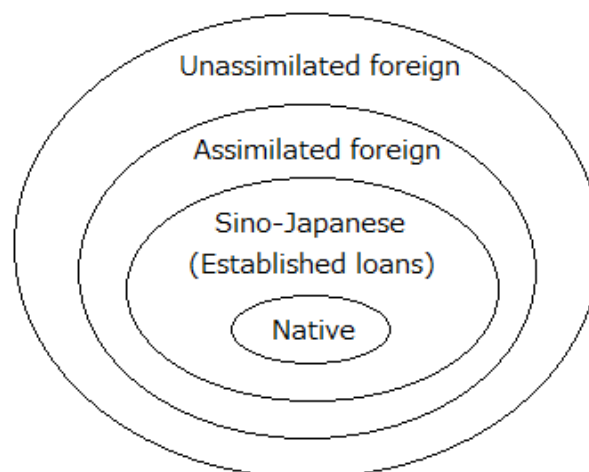


Figure 2.1 Lexical strata in a core-periphery structure adapted from Itô and Mester (1999).

According to Itô and Mester (1999), lexical items in each subset share certain constraints. Additionally, while lexical items in the core stratum (i.e., native words) are strongly constrained

by phonological rules, as a stratum departs from the core stratum, the rules become less and less constrained. That is, general syllable constraints apply to lexical items in all strata, and hence consonant clusters and a final coda are avoided by vowel epenthesis in loanwords. Conversely, the unassimilated foreign stratum is subject to fewer phonological constraints than other lexical strata. Therefore, while obstruent voiced geminates are not allowed in native or Sino-Japanese strata, they are allowed to appear such as *beddo* ‘bed’ in the unassimilated foreign stratum on the periphery. Itô and Mester used the framework of Optimality Theory to argue that the different degrees of phonotactic restriction in different foreign strata are attributed to a higher-ranked constraint, either being a markedness constraint or faithfulness constraint. That is, unassimilated loanwords are enforced by a faithfulness constraint which is ranked higher than a markedness constraint. On the other hand, assimilated loanwords are enforced by a higher-ranked markedness constraint against a faithfulness constraint. Thus, voiced geminates are not allowed in the assimilated foreign stratum, as in *betto* ‘bed’, by devoicing the obstruents, which obeys the markedness constraint. Lexical items such as ‘knob’ and ‘pub’ are resistant to gemination and are borrowed as /nobu/ and /pabu/ that are also assimilated foreign items.

The lexical stratification is not only strongly related to phonological properties but also related to syllabic nature of writing systems in Japanese. “This stratification corresponds in kind to the distinction in English between the Germanic versus the Latinate vocabulary, but is more accessible and conscious to the nonspecialists because of its reflection in the writing system” (Itô & Mester, 1999, p. 63).

Following sections will briefly present relevant information regarding the Japanese phonological system in relation to the lexical stratification discussed in §2.2.1 and §2.2.2. Some important phonotactic constraints that differentiate sublexicon classification are also discussed. Then in §2.2.3, empirical studies of the psychological reality of Japanese lexical stratification are reviewed. Note that Hebon-style Romanisation is used to represent Japanese words, and [u] rather than [ɯ] is used as the phonetic representation of /u/ through this thesis. In this thesis, both the assimilated and unassimilated foreign loans are referred to as loanwords and these strata are called a loanword stratum.

## 2.2 Japanese Phonology

### 2.2.1 Phoneme inventory

Modern Japanese has five vowels /a, e, i, o, u/, and vowel length is contrastive (Akamatsu, 2000; Shibatani, 1990; Tsujimura, 1996; Vance, 2008). As well as vowels, some aspects of the phonology of Japanese consonants are very important for this thesis. The consonantal phonemes of Japanese are presented in Table 2.1; common allophones or consonants that occur restrictedly in loanwords are indicated by parentheses.

Some Japanese consonants vary allophonically, depending on phonological environments. The alveolar consonants /t/, /d/, /s/, /z/ and the glottal fricative /h/ are palatalised when they occur before the high vowel /i/. Alveolar /t/, /d/ and glottal /h/ are also realised as [ts], [dz] and [ϕ], respectively, when they are followed by the high back vowel /u/. A nasal uvular /ɴ/ is called a moraic nasal, when it occurs in coda position. The allophonic relationships are expressed in phonological rule format in (1).

Table 2.1 Consonants of Japanese. Adapted from Akamatsu (2000) and Vance (2008)

	Bilabial	Alveolar	Alveolo-Palatal	Palatal	Velar	Uvular	Glottal
Plosive	p b	t d			k g		
Nasal	m	n			(ŋ)	ɴ	
Fricative	(ϕ)	s z	(ç) (z)	(ç)			h
Affricate		(ts) (dz)	(tç) (dz)				
Approximant			j		ɥ		
Liquid			r				

#### (1) Distribution of Consonants

##### Palatalization

/t/ → [tç] /\_i

##### Examples

/mati/ [matçi] ‘town’

/d/ → [dz] /\_i

/tidimu/ [tçidzimu] ‘shorten’

/s/ → [ç] /\_i

/hasi/ [haçi] ‘edge’

/h/ → [ç] /\_i

/hito/ [çito] ‘human’

##### Labialization

/h/ → [ϕ] /\_u

/huku/ [ϕuku] ‘clothes’

## Affrication

/t/ → [ts] / \_ u

/katu/ [katsu] ‘win’

/d/ → [dz] / \_ u

/tedzukuri/[tedzukuri] ‘handmade’

Thus, there are some co-occurrence restrictions on consonant-vowel sequences of CV syllables. However, while some CV combinations are not allowed in the native phonotactics of Japanese, they are acceptable in nativised loanwords. For example, [ti] and [tu] do not occur in the native Japanese syllable inventory due to a Japanese allophonic rule as noted above (Hirayama, 2003; Irwin, 2011; Kubozono, 2015). On the one hand, [ti] is broadly accepted in contemporary Japanese loanwords, for example, [tii] ‘tea’ and [borantia] ‘volunteer’, while ‘team’ was borrowed as *chiimu* [teiimu] in older loanwords (Pintér, 2015). Another example, [ɛe] does not occur in native vocabularies but it appears in loanwords such as [æeru] ‘shell’. Thus, [seru] ‘cell’ and [æeru] ‘shell’ are contrastive in the loanword stratum. Moreover, as mentioned in (1), while the labial fricative [ɸ] can allophonically occur only before /u/ in native and Sino-Japanese vocabulary, it can appear before any vowel in loanwords, such as [ɸaito] ‘fight’, [ɸirumu] ‘film’, [ɸurii] ‘free’, [kaɸe] ‘café’, and [ɸooku] ‘fork’. Vance (1987) calls traditional allophonic CV sequences such as [tei] and [tsu] the ‘conservative’ variety, while [ti] and [tu] sequences are called the ‘innovative’ variety. CV constraints on modern Japanese result from historical allophonic changes in the Japanese language sound system, as well as resulting from the influence of loanwords (Pintér, 2015). Pintér (2015), for example, claims that the “innovative variety [of Japanese] ... accommodates (almost) all logically possible CV combinations” (p. 125). Such sequences of innovative variety are beginning to establish themselves in Japanese phonology (Kubozono, 2015). Pintér (2015) sees the innovative variety as emergent contrasts, suggesting these forms are not simply contextually predictable allophones.

In addition to the set of basic syllable constraints of Japanese, there is the voiceless labial stop [p] restriction depending on according to the lexical strata, even though the [p] is a licit phoneme in the Japanese language as a whole (Itô & Mester, 1999; Nasu, 2015). On the one hand, in native-Japanese and Sino-Japanese strata, the voiceless labial stop [p] cannot appear as a syllable onset following a vowel. It is tolerated only in a geminated or at least partially geminated form (*kappa* ‘river imp’, *nippoN* ‘Japan’, and *kampai* ‘cheers’, but never \**kapa* or \**nipoN*) (Itô & Mester, 1995, p. 819) That is, in these strata, the voiceless labial stop [p] cannot appear freely as a surface form because underlying singleton /p/ is debuccalised to [h], appearing allophonically in labial [ɸ] before a high back vowel, and in palatal [ç] before a

high front vowel (Itô & Mester, 1999; Nasu, 2015). On the other hand, the voiceless labial stop [p] freely appears as a contrastive surface segment in loanword strata (e.g., *paato* ‘part’ vs. *haato* ‘heart’) (Itô & Mester, 1999; Nasu, 2015).

As well as vowel length, consonant length is contrastive in Japanese phonology, as shown in the minimal pairs below (2). A single consonant is referred to as a singleton, whereas long consonants are referred to as geminates. In general, geminates occur in verb inflection, in compounds, as well as in intensified forms of adverbs and mimetics (Kawagoe, 2015). As for the phonetic property of geminates, closure durations in geminates are more than twice the length than that of singletons in general (Beckman, 1982; Han, 1994; Kawahara, 2015 for summary). In addition, vowels preceding geminates are longer than those preceding singletons, however, vowels followed by geminates are shorter than those of singletons (Han, 1994; Idemaru & Guion, 2008).

- (2) [kata] ‘shoulder’ vs. [katta] ‘won’  
 [haken] ‘dispatch’ vs. [hakken] ‘discovery’  
 [hosa] ‘assistant’ vs. [hossa] ‘attack’

The set of geminate consonants in Japanese is important. In fact, “gemination takes place for various purposes, to remedy phonotactic structure, for intensification, to show the integrity of a compound word, or to attain a certain prosodic structure” (Kawagoe, 2015, p. 98). Essentially, this set is comprised of the voiceless obstruents and they occur only word-medially (Akamatsu, 2000). Secondly, voiced obstruent geminates are not allowed in non-foreign lexicon (i.e., native-Japanese, Sino-Japanese) (Itô & Mester, 1999; Itô, Mester, & Padgett, 1995). As shown in (3), native phonology allows voiceless obstruent geminates but it prohibits voiced geminates. The intensive *-ri* adverb suffix induces gemination of root-final consonants as in (3a). However, when this consonant is a voiced obstruent, gemination are avoided and alternated by nasalization of the first part of the geminates (i.e., a homorganic nasal + voiced obstruent sequence) as in (3b).

- (3)
- |    |             |          |                             |
|----|-------------|----------|-----------------------------|
| a. | biku(-biku) | bikku-ri | ‘surprisingly, frightening’ |
|    | hisō(-ka)   | hisso-ri | ‘secretly’                  |
|    | hono(-ka)   | honno-ri | ‘dimly, faintly’            |

b.	zabu(-zabu)	zambu-ri	*zabbu-ri	‘jumping into water’
	koga(-su)	konga-ri	*kogga-ri	‘toasted, roasted’

(Itô & Mester, 1999)

On the other hand, voiced obstruent geminates are allowed in loanwords such as *baggu* ‘bag’, or *reddo* ‘red’. That is, when foreign words are borrowed into Japanese, word-final obstruents preceded by a lax vowel undergo gemination in words such as the English word *bag* → /baggu/, regardless of voicing types of coda consonants and [u] is epenthesised after the geminate. Voicing in geminates is contrastive in loanwords like *bakku* “back” vs. *baggu* “bag”.<sup>5</sup>

### 2.2.2 Syllable structure

Japanese syllable structure is relatively simple in comparison to that of English, consisting of a consonant-vowel (CV) or vowel (V) (Itô & Mester, 1999; Tsujimura, 1996). Thus, an open syllable is the basic form. Only a syllable-final nasal (e.g., /shimbun/ [ɕimbun] ‘newspaper’) or the first part of a geminate consonant (e.g., /gakko/ [gakkoo] ‘school’, but not [pt], [kt]) can occur in coda position as shown in (2). Of these, only nasals are allowed in the word-final position. The syllable constraints of Japanese are observed in all lexical strata (Itô & Mester, 1999).

Syllable Base		Mora Base	
(4)	[ɕim.buɴ] (CVC.CVC)	‘newspaper’	[ei.m.bu.ɴ] (CV.C.CV.C) しんぶん
	[gak.koo] (CVC.CVV)	‘school’	[ga.k.ko.o] (CV.C.CV.V) がっこう

In general, Japanese speakers divide words into morae. A mora consists of either a vowel (V) or a vowel preceded by a consonant (CV). For example, /ga/ is both monomoraic and monosyllabic, and is thus called light syllable. The words such as /ɕimbun/ and /gakko/ contain heavy syllables; monosyllabic but bimoraic. Japanese has four types of heavy syllables: first half of geminate, moraic nasal, second part of long vowel and second part of diphthong (Kawahara, 2016). Geminate consonants are not allowed to occur after a long vowel or

---

<sup>5</sup> But also voiced obstruents in the coda position in English words optionally undergo devoicing when they co-occur with another voiced obstruent, for example, /baggu~/bakku/ ‘bag’ (Kawahara, 2006, 2011; Nishimura, 2003). Thus, /bakku/ ‘bag’ and /bakku/ ‘back’ become homophonic and indistinguishable from each other (Kubozono, 2015). The devoiced pronunciation such as /bakku/, /betto/ are treated as “undesirable” in a Japanese accent dictionary (Kindaichi & Akinaga, 2014, p28).



diphthong, which means that superheavy syllables (i.e., trimoraic syllables) are restricted (Kubozono et al., 2008).

Last but far from least, the Japanese writing system is strongly related to phonotactics, syllable structure and the lexical strata. The writing system of modern Japanese combines three different scripts: phonographic *hiragana* and *katakana* with logographic *kanji* adapted from Chinese characters. The *hiragana* and *katakana* are a syllabary; each letter corresponds to one mora. It can be seen that the word *shimbun* ‘newspaper’ and *gakko* ‘school’ in (4) are four segmented morae that conform to the four letters of *kana* used in the written form. In addition, different lexical classes are written with different sets of letters. While *hiragana* is used to write particles, grammatical inflections and native words, *katakana* is basically used for loanwords and mimetic words. *Kanji* is used for Sino-Japanese words, but native words are also written in *kanji*. Thus, Japanese students need to learn stratal affiliations of each lexical item.

In this section, the phonology of Japanese in relation to the lexical stratification of Japanese was reviewed. Although each stratum includes some internal phonological similarity in the Japanese language as a whole, there are also some distinctive phonological properties. In terms of the loanword stratum, the phonotactic requirements are less restrictive than in native and Sino-Japanese strata. Firstly, nativised loanwords frequently contain innovative CV sequences which do not appear in native-Japanese and Sino-Japanese words (Irwin, 2011; Kubozono, 2015; Pintér, 2015; Vance, 2008). Secondly, the voiceless labial stop [p] can freely appear as a licit surface form (Itô & Mester, 1995, 1999; Nasu, 2015; Shibatani, 1990). Third, voiced obstruent geminates are allowed while they do not occur in the native and Sino-Japanese strata. In addition to the phonotactics, Japanese loanwords are signalled by way of a different set of orthography (*katakana*) from native-Japanese and Sino-Japanese. The following subsection discusses findings from empirical studies on lexical strata in Japanese and their implications for the psychological reality of lexical strata in Japanese.

### **2.2.3 Psychological reality of lexical strata in Japanese**

Some auditory perception studies have examined the psychological reality of lexical stratification (Gelbart & Kawahara, 2007) and sublexicon phonotactics (Moreton & Amano, 1999). Lexical stratification is very salient for adult speakers of Japanese who have accessed knowledge of lexical stratification in auditory perception (Gelbart & Kawahara, 2007; Moreton & Amano, 1999). Listeners’ perceptual boundaries are affected by stimulus items which contain different lexical stratum cues.

Moreton and Amano (1999) investigated the stratum-specific phonotactic effect on perception of vowel length in nonwords of the form [C<sub>1</sub>oC<sub>2</sub>a(a)]. The final [a] was an ambiguous segment, varying in duration as a short-long continuum. While vowel length is contrastive in Japanese, word-final [aa] is only found in loanwords. Therefore, the positional phoneme probability of [aa] is high in loanwords but nil in other lexical strata. Moreton and Amano consider static distribution patterns of phonemes in each lexical stratum based on a corpus. Then, they selected C<sub>1</sub> and C<sub>2</sub> to provide stratum cues in stimuli; [rj] and [hj] for Sino-Japanese, [p] and [ϕ] for foreign stratum, [r] and [t] for neutral contexts, resulting in nine possible combinations of C<sub>1</sub> and C<sub>2</sub> in stimuli. While these palatalised consonants dominantly appear in Sino-Japanese words, the labial consonants freely appear in loanwords but not in the Sino-Japanese stratum. Moreton and Amano hypothesised that stratum phonotactics would create boundary shift since the Sino-Japanese words were expected to lack [aa]. That is, when [rj] and/or [hj] occurred with [aa], ambiguous segment [aa] needs to have an acoustically longer duration to be perceived as [aa] by listeners, than when [p] and/or [ϕ] occurred with [aa]. In order to test this, 24 L1 Japanese speakers were asked to judge whether the stimuli they heard were long or short vowels by clicking buttons on a screen. The buttons were labelled with stimulus words written in *katakana*. Results showed that as expected, boundary perception was shifted according to consonantal cues to stratal affiliation, even when the triggering phoneme (i.e., C<sub>1</sub>) was not immediately adjacent to the ambiguous segment. In other words, listeners were more likely to judge nonwords as [CoCaa] when given foreign cues than in contexts lacking foreign cues, if the duration of [aa] is the same. The study shows that perception can be affected by lexical stratum phonotactics. Thus, the psycholinguistic reality of stratum-specific phonotactics for Japanese L1 speakers was supported. Moreton and Amano concluded that simple segment-to-segment transitional probabilities cannot explain the findings as the trigger phoneme (i.e., C<sub>1</sub>) was at a distance of three phonemes from the ambiguous segment. Although a transitional probability was not available for the listeners, a co-occurrence probability was available. The results indicate that native Japanese speakers are sensitive to conditional probabilities by which phonemic features are likely to co-occur in sublexicons. In addition to the effect of C<sub>1</sub>, the effect of C<sub>2</sub> was also found. When two Sino-Japanese cues are available, statistical probability of classification into the Sino-Japanese word would be higher than contexts presenting only one cue. This suggests that listeners are not only sensitive to the characteristics of the input, but also to the cumulative phonotactic probability. This study suggests that Japanese L1 speakers exploit different kinds of probabilistic information to associate the lexical cues and the ambiguous segment of the input.

Gelbart and Kawahara (2007) also found a stratum-specific biasing effect on speech perception by adult Japanese L1 speakers. In this study, instead of using nonwords, existing native and foreign (i.e., loanwords) words were selected, and eight word pairs were created. Each pair consisted of one native and one loanword that had the same accent position. While these words did not contain specific phonotactic cues to stratal affiliation, “the lexical item itself was the cue” (Gelbart & Kawahara, 2007, p. 64), such as *nasa* ‘NASA’ and *kurabu* ‘dance club’ for foreign words, and *mosa* ‘tough guy’ and *narabu* ‘line up’ for native words. Following the study in Moreton and Amano (1999), the stimuli contained length contrasts [a]~[aa] in the word-final position, but also the voiced stop geminacy contrasts [b]~[bb], [d]~[dd], and [g]~[gg] (e.g., *kurabu* vs. *kurabbu*) as ambiguous segments with differing lengths, yielding a continuum of stimuli. Voiced geminates are inhibited in native words as well as the word-final [aa]. Therefore, a stratum-specific biasing effect was expected for voiced geminates; Japanese listeners’ categorisation would be biased toward geminate consonants, when listeners perceive that the stimuli belong to the foreign stratum in which geminates are phonotactically legal. Twenty-six Japanese L1 speakers were asked to judge whether stimuli they heard were the standard form (singleton consonants, [a]) or long form (geminates, [aa]). Consistent with Moreton and Amano, Japanese listeners were more likely to categorise the final vowels as long [aa] in the foreign stimuli than in the native stimuli. As for the obstruent continua, pairs that had obstruent continua in third consonants also showed boundary shift. However, pairs that had the obstruent continuum in second consonants did not show boundary shift. Contrary to the prediction, listeners’ categorisation was biased toward voiced geminates in native words than in foreign words. Gelbart and Kawahara attributed this finding to emphatically geminated voiced geminates. That is, when words are pronounced emphatically, onset of the second syllable in native word undergoes geminates (e.g., *sugoi* ‘very’, *suggoi* ‘very (emphatic)’). Another possibility is that the unexpected findings might also be related to statistical information contained in phoneme sequences of words. That is, some phoneme sequences might be more common in a lexical stratum than in another stratum. For example, phoneme sequences of *nega* ‘negative’ as a foreign stimulus in the study occur in native words such as *negai* ‘wish’ or *negaeri* ‘roll-over’. Therefore, words containing such sequences might have influenced perception, causing perceptual bias toward single obstruent rather than geminates. One way or another, adult Japanese L1 speakers are sensitive to lexical stratification.

Experimental evidence suggests that Japanese L1 speakers have knowledge not only of entire lexicons but also phonological sublexicons, detecting the phonotactic patterns of words

according to their lexical affiliation. The question that this thesis investigates is whether L2 learners are aware of this stratification, and if such knowledge can be acquired by L2 learners.

Recently, Morita (2018) showed that stratal affiliation of Japanese lexicons are learnable from phonotactics. His study was grounded in a Bayesian learning-based computational clustering model that was applied to Japanese and English words from corpora, which is able to predict etymological lexical subclass from segmental phonotactics. As for English lexicons, Morita tested etymological classification of Germanic and Latinate lexicons. Specifically for Japanese, the computational model learned a substantial number of nouns (30,554 type frequency words) from a corpus as Japanese nouns are not inflected like verbs or adjectives. Morita proved not only the coexisting of different subphonological systems within the two languages but also the learnability of sublexicons from naturalistic data. In addition, his sublexicon learners were applied to the previous study of the psychological reality of sublexicon phonotactics, capturing the quantitative patterns of the experimental results. Importantly, this study suggests, whether speakers be native or L2 learners of Japanese, users of Japanese language should be capable of learning such sublexicon-specific knowledge by using the statistical language learning mechanisms from their accumulated lexicons.

The next section will focus on Japanese loanword phonology. Certain phonological processes apply only to loanwords. Epenthetic vowels are a phenomenon existing only within the loanword lexicon. An interesting aspect of the process of epenthesis is accompanied by the process of obstruent consonant lengthening following the lax vowels (i.e., gemination). Within loanwords, voiced geminates are less likely than voiceless geminates. In addition, there are stochastic patterns, such that the likelihood of different voiced geminates varies. After a short review for general loanword phonology, existing studies on loanword phonology focusing on epenthesis and gemination are surveyed.

## **2.3 Loanword Phonology**

This section will first outline general loanword phonology. Then, a history of loanwords in Japanese is briefly explained in §2.3.2. I then review two specific strategies that Japanese speakers use to adapt foreign words into the Japanese lexicon in §2.3.3.

### **2.3.1 Introduction**

Loanwords are words in a language that are borrowed from another language (Kang, 2011). Loanword adaptation refers to the process by which words are altered, when they fail to meet

the phonological requirements of the host language (Kang, 2011). Loanword phonology is the study of the function and organisation of sounds in words that have been borrowed, and it is a subcomponent of the native language phonology in a language. That is, while loanword phonology is constrained by native phonology to some extent, specific rules are applied to only loanwords but not to native lexicons. Many languages show that loanwords incorporate novel features that are not allowed in native lexemes, different alteration strategies and prosodic features, and even occasionally showing ‘unnecessary’ adaptations (e.g., Kang, 2003; Kenstowicz & Suchato, 2006; Kubozono, 2015; Peperkamp, 2004). On the other hand, these differences clarify the nature of phonology in host languages and the phonological knowledge of a first language.

In terms of loanword phonology, studies in various languages show that loanword adaptation generally involves phoneme substitution, and systematic modification of non-native sequences through aspects such as neutralisation, prosthesis, metathesis, epenthesis, deletion, and gemination (e.g., Davidson, 2006; Fleischhacker, 2001; N. Hall, 2011; Kang, 2011; Kawahara, 2011; Kenstowicz, 2007; Miao, 2005; Paradis & LaCharité, 1996; Peperkamp, 2004; Shirai, 2012; Uffmann, 2006). On the segmental level, in general, host languages borrow phonemes based upon how phonetically close they are to counterparts in the host language (Hock & Joseph, 1996; Kubozono, 2015). For example, Japanese does not have the same distinction between /ɪ/ and /l/ as English does, hence these sounds are neutralised as /ɪ/ ([ɾ]) and become homophonous. On the phonotactic level, illicit sequences that do not comply with native phonotactics are generally modified. For example, the English word ‘ski’ is commonly pronounced as [iski] in Egyptian Arabic by adding a vowel before the first consonant; a process termed prosthesis (Broselow, 1987 cited in Fleischhacker, 2001). In another case, English loanwords in Māori language can be spoken with an extra vowel inserted to break up consonant sequences. Furthermore, the vowel inserted is often the same quality as the vowel in the adjacent syllable (i.e., copy vowel epenthesis). For example, the English word ‘blue’ [blu:] is adapted as *puruu*, while another English word ‘ink’ [ɪŋk] is adapted as *ingiki* (Kearns, 1990). Thus, modification strategies are dependent on the languages and the medium through which words are adapted (e.g., orthographic input or audio input).

### **2.3.2 Loanwords in Japanese**

As mentioned in the previous section, the Japanese lexicon consists of four strata in terms of etymology: native Japanese, Sino-Japanese (old loans from Chinese), assimilated foreign (older and more nativised loans of non-Chinese origin) and unassimilated foreign (newer and less nativised loans of non-Chinese origin) (Itô & Mester, 1995, 1999, 2001). Japanese has abundance of loanwords and a long history of borrowing words from other languages. From 1639 to 1853, Japan had an isolationist foreign policy, and movement in and out of Japan was strictly controlled. As a consequence of contact with permitted Portuguese and Spanish missionaries or Dutch traders during the period, lexical borrowing from these languages into Japanese occurred (Irwin, 2011). After the policy ended, trade and diplomatic relations with the United States, United Kingdom, Russia, Netherlands, France, and other countries have facilitated access to factors such as European technology, science, philosophy, and culture (Kay, 1995). The majority of loanwords were derived from English by the turn of the 20th century, and this tendency has not changed up to the present day (Irwin, 2011). Loanwords, especially those from English vocabulary, are used on a daily basis in publications, and are also perceived through media such as TV and radio. According to a loanword survey conducted by the National Institute for Japanese Language and Linguistics (NINJAL, 2005), loanwords used in magazines tripled from 1956 (9.8%) to 1994 (34.8%) as token frequency. In relation to phonological aspects, loanwords (except from Chinese) are written in the distinctive Japanese phonetic script, the *katakana* syllabary. This unique situation exhibits certain phonological processes that apply only to loanwords. In the next section, I will discuss two specific strategies that are used to adapt foreign words into the Japanese lexicon. These nativisation processes are important for the hypotheses investigated by this study.

#### **2.3.2.1 Vowel epenthesis**

When languages borrow foreign words, some segments undergo sound changes in order to comply to the phonotactics of the host language. As Moreton (2002) mentioned, “[t]he alternations induced by phonotactics are categorical rather than gradient, and systematic rather than arbitrary” (p. 5).

As for the segmental correspondence, since the five Japanese vowels have length contrasts, lax vowels in the source words are borrowed as Japanese short vowels in Table 2.2. On the other hand, tense vowels are borrowed as Japanese long vowels (Kubozono, 2015).

Table 2.2 Basic correspondence between lax English vowels and their Japanese counterparts, Kubozono (2015, p. 315)

English vowel	Japanese vowel	Examples
[i]	/i/	pin, kiss, milk
[e]	/e/	pen, test, best
[ʌ]	/a/	but, cut, mother
[æ]	/a/	bat, back, lag
[u]	/u/	push, bull, put
[ɔ]	/o/	box, pot, dog

Similar to vowels, consonants in the source languages also turned into phonetically closest consonants in Japanese. For example, Japanese does not have the dental fricative /θ/, hence the alveolar fricative /s/ is substituted even though their places of articulation are different. Although the adaptation process of onset consonants is relatively straightforward, coda consonants in Table 2.3 are necessary to be adapted as open syllables because syllable-final consonants are not allowed except moraic nasals in Japanese.

In many languages, vowel epenthesis is a common repair strategy for coda consonants and consonant clusters from the source language that do not meet the phonology of the host language (Fleischhacker, 2001; N. Hall, 2011; Kang, 2011; Uffmann, 2006). Vowel epenthesis refers to an additional vowel in utterance (N. Hall, 2011), and it is the most common syllable modification strategy (Kang, 2011; Weinberger, 1994). As is the case with other languages, vowel epenthesis is employed in the Japanese language as a syllable modification strategy (Hirayama, 2003; Kubozono, 2015), because Japanese basic syllable structure is CV (Tsujimura, 1996). In borrowings, the consonantal codas and consonant clusters undergo the process of vowel epenthesis, by which these illicit segments in the source language change into licit open syllables in Japanese. For example, the English word ‘pipe’ [paɪp] is commonly pronounced as [paɪpu], with [u] occurring in word-final position since consonants other than [N] do not occur word-finally in Japanese (Kubozono, 2015). Vowel epenthesis serves to make non-native structures more native-like (e.g., Hirayama, 2003; Itô, 1989; Kubozono, 2015; Smith, 2006).

Table 2.3 Basic correspondence of coda consonants, Kubozono (2015, p. 322)

English	Japanese	Examples
[p]	[p]	cap, top
[t]	[t]	pot, mat
[k]	[k]	pack, duck
[b]	[b]	cab, pub
[v]	[b]	love, live
[d]	[d]	bed, god
[g]	[g]	bag, big
[f]	[ϕ]	tough, knife
[θ]	[s]	bath, mouth
[s]	[s]	bus, mouse
[ʃ]	[ʃ]	push, smash
[ð]	[z]	clothes, bathe
[z]	[z]	nose, pose
[tʃ]	[tʃ]	beach, peach
[dʒ]	[dʒ]	bridge, judge
[m]	[m]	ham, gum,
[m]	[m] (moraic nasal)	camp, hamburger
[n]	[n] (moraic nasal)	ten, tent, tender
[ŋ]	[ŋg]	gang, long
[l]	[r] ([r])	pill, pool

Existing nativised loanwords studies agree that there are three epenthetic vowels [i, o, u], depending on the quality of preceding consonant (e.g., Hirayama, 2003; Katayama, 1998; Kubozono, 2001; Kubozono, 2015; Lovins, 1975; Otaki, 2012). The type selected reflects co-occurrence restrictions on CV sequences in syllables. In the majority of the preceding consonantal contexts, [u] is selected and is generally considered to be the default epenthetic vowel. The high back [u] appears to be the least salient vowel in the Japanese vowel inventory as it is the shortest vowel and the most susceptible to weakening and deletion in Japanese (Hirayama, 2003; Kubozono, 2015; Sagisaka & Tokuhara, 1984 as cited in Irwin, 2011; Shoji & Shoji, 2014). These characteristics are consistent with the view that the epenthetic vowel is perceptually the least salient in the language (Byarushengo, 1976; Fleischhacker, 2001; Kang, 2003; Kenstowicz, 2007; Shinohara, 1997; Steriade, 2001a, 2008). The high front vowel [i] is inserted after the palato-alveolar affricates [tʃ], [dʒ], and the voiceless velar [k]. Early



loanwords show an epenthetic [i] after the voiceless velar [k], whereas recent loanwords exhibit an epenthetic [u]. Some loanwords with an epenthetic [i] after [k] have doublet forms with an epenthetic [u]. For example, with the English word, *text* [tekst], the Japanese borrowing is [te.ki.su.to] or [te.ku.su.to] (Irwin, 2011). [i] insertion occurs after [tʃ] and [dʒ], and the front vowel [i] shares similar articulatory and perceptual properties with these consonants (Hirayama, 2003; Kubozono, 2015). In addition to [i] and [u], the mid back vowel [o] typically occurs after the alveolar stops [t, d]. The reason for the insertion of [o] after alveolar stops is that [tu], [du], [ti], and [di] are phonotactically licit in the native Japanese syllable inventory. The choice of [o] also seems to be associated with perceptual properties, and while the original consonants are preserved with inserting [o] after alveolar stops, inserting [u] after alveolar stops could be realised as affricates [ts] and [dz] due to an allophonic rule in Japanese (Hirayama, 2003; Irwin, 2011; Kubozono, 2015). The phonological rules noted above are formulated in (5).

(5) Epenthetic vowels

- (i)  $\emptyset \rightarrow i / tʃ, dʒ \_ \#$  and  $k \_ \#$
- (ii)  $\emptyset \rightarrow o / t, d \_ \#$
- (iii)  $\emptyset \rightarrow u /$  in all other contexts and  $k \_ \#$

(Irwin 2011; Shoji & Shoji, 2014)

For example, the English word *Christchurch* will include all possible epenthetic vowels according to the preceding consonantal contexts, in order to modify the word structure to English loanwords in Japanese (i.e., *Christchurch* /kɹɪstʃɜːtʃ/ → [kuraɪsutotʃaatʃi]). The first consonant cluster is broken up by adding the high back vowel [u] after [k], and the second cluster is repaired by adding [u] after [s] and [o] after [t], respectively. The single consonant in word-final position is repaired through inserting the contextual appropriate epenthetic vowel [i]. Thus, the process of epenthesis is productive, and the choice of vowels is mostly predictable. Therefore, the choice of the epenthetic vowels is categorical generalisation rather than gradient generalisation (e.g., Ernestus, 2011). Note that this dissertation considers only (ii) and (iii) in (5) for later experiments.

The epenthetic vowel [u] has a higher frequency as it is used after 10 coda consonants /p, b, k, g, ɸ, s, ʃ, z, m, ɾ/ in Table 2.3, in comparison to the vowel [o] which occurs only after alveolar stops /t, d/. Thus, it would be possible that language users overgeneralise an epenthetic rule in which [u] can be used in any contexts. As discussed in §1.3, with frequent exposure to specific contexts, adult learners restrict generalisation. In order to extract the patterns of

epenthetic vowels from instances of loanwords, generalizing patterns to novel instances, larger lexicons are crucial. This is because less frequent items are more likely to occur in a larger lexicon than in a small lexicon (Frisch et al., 2001).

Interestingly, although [tu] is not as common as another innovation variety [ti], it is not actually illegal in the loanword stratum. That is, [tu] is an attested sequence in the loanword lexical stratum since it is recently acceptable for borrowings such as *tatuu* ‘tattoo’. This is an example of a lexical item in the unassimilated foreign stratum based on the Core-Periphery Structure model (Itô & Mester, 1999) as discussed in §2.1. Thus, the process of epenthesis is greatly constrained by Japanese native phonology.

In summary, epenthetic vowels are a phenomenon only existing within the loanword lexicon, and there are categorical constraints dictating which vowels should be used.

Another interesting aspect of the process of epenthesis is accompanied by the process of obstruent consonant lengthening following the lax vowels (i.e., gemination). Consonant gemination in nativised loanwords will be discussed in the next section.

### 2.3.2.2 Consonant gemination in loanwords

In addition to vowel epenthesis, germination can often be seen in adapting non-native sounds into Japanese. Word-final obstruent consonants preceded by a lax vowel are often borrowed as geminates (e.g., ‘pet’ is borrowed *petto*) (Itô et al., 2017; Kaneko & Iverson, 2009; Katayama, 1998; Koo & Homma, 1989; Kubozono, 2001; Lovins, 1975; Shirai, 1999). As mentioned earlier in this thesis, geminates occur in Japanese phonology but only for voiceless obstruents. The conditions of gemination are asymmetry between voiced and voiceless obstruents in nativised loanwords (Hirayama, 2005; Kubozono et al., 2008). The environment where consonantal gemination occurs is more or less predictable and a systematic sound adaptation pattern exists. The occurrence of gemination depends on the phonological structure of the source words. When English words contain a single voiceless stop /p, t, k/ in word-final position preceded by a lax vowel, those singletons are typically adapted as geminates /pp, tt, kk/ in Japanese loanword phonology (Koo & Homma, 1989; Kubozono, 2001; Lovins, 1975; Shirai, 1999), as shown in (6).

(6)	English	Japanese
	cup	ka[pp]u
	hip	hi[pp]u

pet	pe[tt]o
kit	ki[tt]o
book	bu[kk]u
neck	ne[kk]u

This pattern of adaptation is well-attested by nativised loanwords. Shirai (1999) investigated the patterns of gemination with relation to the phonological environments of the source words, using a loanword dictionary containing 12,000 words from English, German and other languages. She found that 492 out of 3,999 loanwords derived from English have gemination. The most commonly occurring condition is the context mentioned above, where geminated consonants appear after a lax vowel (333 out of 492). In this context, over 90% of voiceless stops become geminates: [p] 100%, [t] 92.3%, [k] 98%. On the other hand, when the same voiceless stops are preceded by a tense vowel, they are not geminated (i.e., remaining as singletons).

The structure of source words and typical adaptations are as shown in (7). For example, the English word ‘pet’ [pet] is commonly pronounced and written as [petto], with [o] occurring in word-final position since a single consonant cannot stand alone in this position and the phonological rule (5-ii) applied. The occurrence of [t] word-medially follows the word-final gemination rule in (7). As a result, ペット [pet.to] is a loanword adapted from English to Japanese, which conforms to loanword phonology in Japanese.

(7) Word-final germination

$$\emptyset \rightarrow C_1 / V_{[\text{lax}]} C_{1[-\text{son}, -\text{voice}]} \_ \# \quad (\text{cf. Shirai, 1999, p. 1})$$

Gemination is exhibited by native speakers of Japanese in empirical studies of loanword adaptation in which real English real words and nonwords were used. In both writing (Preston & Yamagata, 2004) and oral production (Kaneko & Iverson, 2009), the final voiceless stop consonants were geminated after lax vowels almost 100% of the time in both studies. Similarly, for Takagi and Mann (1994), the results of four-alternative forced choice tasks showed that for nonwords with the CVC structure, the geminate modification (CVCCV) was more likely to be selected as the best representation of the target words for the lax vowels, while vowel lengthening (CVVCV) was utilised for tense vowels.

In terms of voiced segments, despite disallowing geminates of voiced obstruents in the native phonology of Japanese, voiced geminates are allowed in loanwords. However, there is a bias against voiced geminates in Japanese. While the occurrence of voiceless geminates is predictable and stable, occurrence of voiced geminates is not stable and unpredictable even under the same segmental and contextual conditions in the source words. While voiced stops are geminated in (8a), they fail to geminate, as in (8b).

(8a)

<u>English</u>	<u>Japanese</u>
snob	sno[bb]u
web	we[bb]u/webu
kid	ki[dd]o
pad	pa[dd]o
pig	pi[gg]u
wig	ui[gg]u
tag	ta[gg]u/tagu

(8b)

pub	pabu
tub	tabu
mug	magu

According to Hirayama (2005) the voiced stop geminates in nativised loanwords are unstable and asymmetrical because of the gemination rates of different places of articulation in word-final position. Hirayama shows gemination rates of three places of articulation in word-final position based on three distinctive surveys of nativised loanwords (Hirayama, 2005; Maruta, 2001 cited in Kawagoe & Arai, 2002; Shirai, 1999); the labial [b] rarely undergoes gemination at 11–23%. The coronal [d] is more apt to gemination with 58–83%, and the velar [g] is around half, 42–55%. Thus, while [dd] quite frequently appears in loanwords, [bb] is rare and [gg] falls between these two (i.e., [dd] > [gg] > [bb]).

Thus, within other sublexicons (i.e., native and Sino-Japanese), there is a categorical constraint preventing geminates. However, within our target sublexicon (i.e., loanwords), there are stochastic patterns, such that the likelihood of different voiced geminates varies. Moreover, voiced geminates are less likely than voiceless geminates in this context.

Together, the adaptation of word-final consonants by epenthesis and gemination will provide a test case for studying the statistical learning of sublexicon phonology on well-formedness judgements. Well-formedness ratings should reflect the probabilistic patterns of loanwords. It is assumed that while the process of epenthesis is categorical and involves rule learning, the process of gemination is gradient. Precisely speaking, the relationship between well-formedness and the likelihood of different voiced geminates would be predicted to be gradient. As discussed in §1.3, both processes use the same learning mechanism. I assume that categorical rules are more readily learned than gradient ones. This is because in order to detect gradient patterns, generally more data is required which entails greater lexical knowledge. Also, some empirical studies show that adult learners produce categorical sounds more accurately than gradient allophonic alternation (Shea & Curtin, 2011).

The following subsections discuss a statistical analysis of data using a corpus showing the pattern of geminates in loanwords, consistent with previous studies. The estimated vocabulary size of native Japanese speakers and learners of Japanese will be also discussed.

### **2.3.2.3 Corpus work for consonant gemination**

The aim of this section is to provide a statistical analysis of data from a large-scale corpus, reflecting actual language use of loanwords in Japanese. I will provide token frequency and lexical representation of nativised loanwords in Japanese in relation to loanwords containing target phonological regularities. The Balanced Corpus of Contemporary Written Japanese (henceforth BCCWJ, National Institute for Japanese Language and Linguistics, 2011) was used, containing written language and etymological information. The written corpus was useful to examine the lexical frequency and estimate listeners' loanword vocabulary size in the present study. It is the largest corpus of the written language used in modern Japanese, containing 185,136 words in types and 104 million words in tokens, that were extracted randomly from such sources as books, magazines, newspapers, business reports, textbooks but also web resources. Without including functional words and proper nouns, the percentages of native Japanese words are 32.98%, for Sino-Japanese (Chinese origin) 43.59%, for loanwords 18.6% and for mixed sources 4.83% (National Institute for Japanese Language and Linguistics, 2011). In the corpus, 21,621 words were tagged as loanwords, which are written in katakana syllabary along with source words, of which 4,846 words without sub-lemma (i.e., source words) were removed from the current data analyses. Almost of all them were pseudo-loanwords (e.g., the word *minikaa* refers to 'microcar') or truncated (e.g., the word *masukomi* refers to 'mass communication'). As a result, 16,771 loanwords remain. Taking previous studies of corpora on

loanwords into consideration, Shirai (1999) examined a 3,999-loanword corpus in Japanese. Thus, the current study provides a wide range of loanwords for analysing the observed tendency described in the previous section.

Firstly, distributions of geminates in the BCCWJ were examined. Focusing on loanwords with geminates using R (R Core Team, 2018), there are 2,224 words with geminates. Narrowing down a target to words that have a lax vowel + stop consonant following contextually appropriate epenthetic vowels, 1,227 words remain such *toppu* ‘top’, *netto* ‘net’, *bokkusu* ‘box’ in Japanese. That is, words like *happii* ‘happy’ and *pakking* ‘packing’ in Japanese, respectively, are not included. The frequency of geminate stops in loanwords at any position occurs in the following order: kk (#503) > tt (#407) > pp (#153) > dd (#129) > gg (#33) > bb (#2). As discussed above, voiced geminate stops are not as common as voiceless stops.

Taking a closer look at loanwords with a target C in word-final position in the source language, voiceless stops /p, t, k/ undergo gemination almost all the time under the expected contexts as shown in Figure 2.2. On the other hand, for voiced stops, gemination is not stable. Alveolar [d] is more likely to undergo gemination than velar [g], and labial [b] is the least likely. Interestingly, while only two English loanwords *snobbu* ‘snob’ and *mobbu* ‘mob’ were geminated, similar loanwords such as /nobu/ ‘knob’ and /stabu/ ‘stub’ do not take a form of geminates. In summary, the gemination of word-final stops occur 95% in the voiceless context while the percentages of each voiced geminates are 80% for [dd], 47% for [gg], and 5% for [bb].

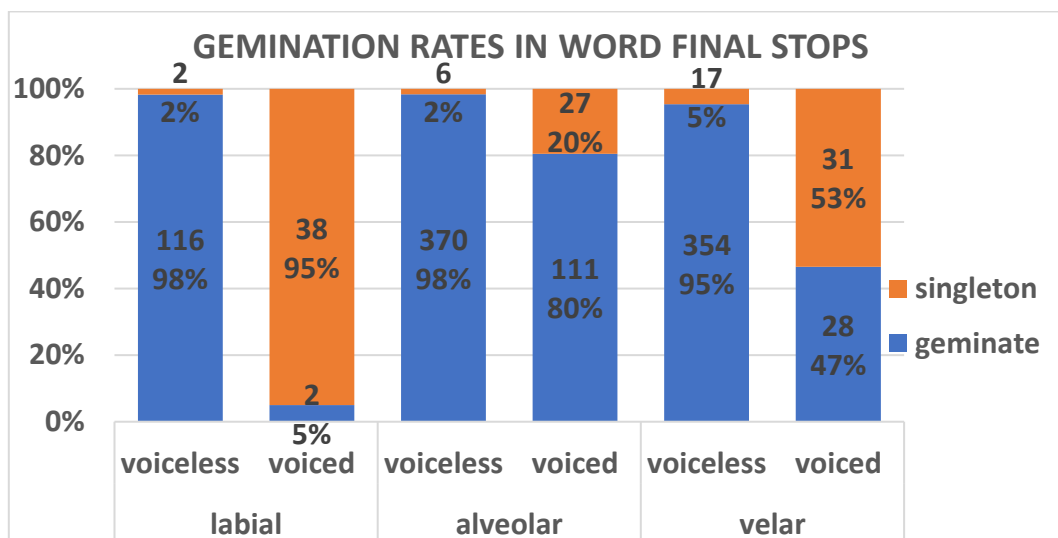


Figure 2.2 Frequency of geminates and singletons for source words with a word-final stop following a lax vowel in the BCCWJ. Each number inside the bars stands for the number of tokens according to voicing types across places of articulation.

Thus, overall the findings from the dataset are consistent with previous studies reported by Hirayama (2005), in which voiced geminates do not occur as frequently as voiceless geminates. The gemination rate from higher to low is [dd] > [gg] > [bb]. The familiar English loanwords such as “red”, “head”, “bed”, “pad”, and “wood” are all realised as gemination, never singleton. The percentage of gemination for voiceless stops is more than 95%.

Given the discussion above, this informs the question as to whether learners are able to detect target regularities from loanwords with geminate stops as related to the size of their loanword vocabulary. In the next section, the size of loanword vocabulary will be estimated for both native speakers of Japanese and L2 learners of Japanese.

### **2.3.3 Frequency of the loanword lexicon and size of the loanword lexicon**

#### **2.3.3.1 Japanese vocabulary size of native Japanese speakers and learners of Japanese**

As we discussed in §1.2.2, there is a strong relation between the size of one’s vocabularies and statistical learning during first language acquisition (Edwards et al., 2004; Graf Estes et al., 2011). Therefore, it would be expected that the size of vocabulary plays a role in statistical learning in L2 acquisition as well as in L1 acquisition. The average vocabulary size of university students in Japan is generally around 40,000 words (Sato, Tajima, Hashimoto, Matsushita, & Sasao, 2017). To obtain this data, Sato et al. (2017) used a 50,000-word vocabulary size test based on written-corpus frequency data to test first-year Japanese university students upon matriculation. Of the 400 people analysed, 96.8% of the students were estimated to have above a 30,000-word-level vocabulary size and 74.5% were estimated to have more than 40,001 words. Surprisingly, some students had an extremely small vocabulary size of around 18,000 words. On the other hand, according to programme information about Japanese courses on the University of Canterbury website, when learners complete an elementary Japanese course (i.e., first year of an undergraduate degree), they are able to understand approximately 300 Japanese words and phrases in addition to the hiragana/katakana syllabary. It is anticipated that learners will understand 800 words and phrases in the subsequent course and 2,500 words and phrases in the intermediate course after that. There are five courses from a beginner level to an advanced level for three years, and the vocabulary size of English-speaking learners of Japanese is estimated at around 3,500 words after completing all courses.

According to Tabata-Sandom (2015), university students taking Japanese in an L2 course rarely reach the vocabulary level at which they can understand more 6,000 words by the end of their study. Thus, although learners’ approximate vocabulary size is estimated by these

course objectives, it seems that the Japanese vocabulary size of L2 speakers is about one tenth of that of the L1 speakers of Japanese.

### 2.3.3.2 Estimated loanword vocabulary size of native Japanese speakers and learners of Japanese

From the dataset, an estimated loanword vocabulary size in native Japanese speakers and learners of Japanese was extracted. In order to do this, I used word frequency rankings in the corpus, which show that each single word has a ranking according to its frequency per 1 million words. For example, for the first 1000 highly frequent words, only 17 loanwords appeared. I calculated how many loanwords were ranked every 1000 most common words and the cumulated total was obtained every 1000 rankings in Figure 2.3. The fact that it is almost linear shows that the loanwords vocabulary is distributed across word frequencies in a manner that closely resemble the non-loanword vocabulary. According to the data, native speakers of Japanese are likely to have 4,367 loanwords if they have a vocabulary of around 39,000 to 40,000 items. As for learners, they tend to have only 295 loanwords, if their vocabulary consists of around 3,000 to 4,000 Japanese words. Thus, the estimated loanword vocabulary size of advanced learners of Japanese is around 300 words, which is closer to the vocabulary size (303 words) in which infants tend to successfully learn phototactically licit stimuli in their native language (Graf Estes et al., 2011).

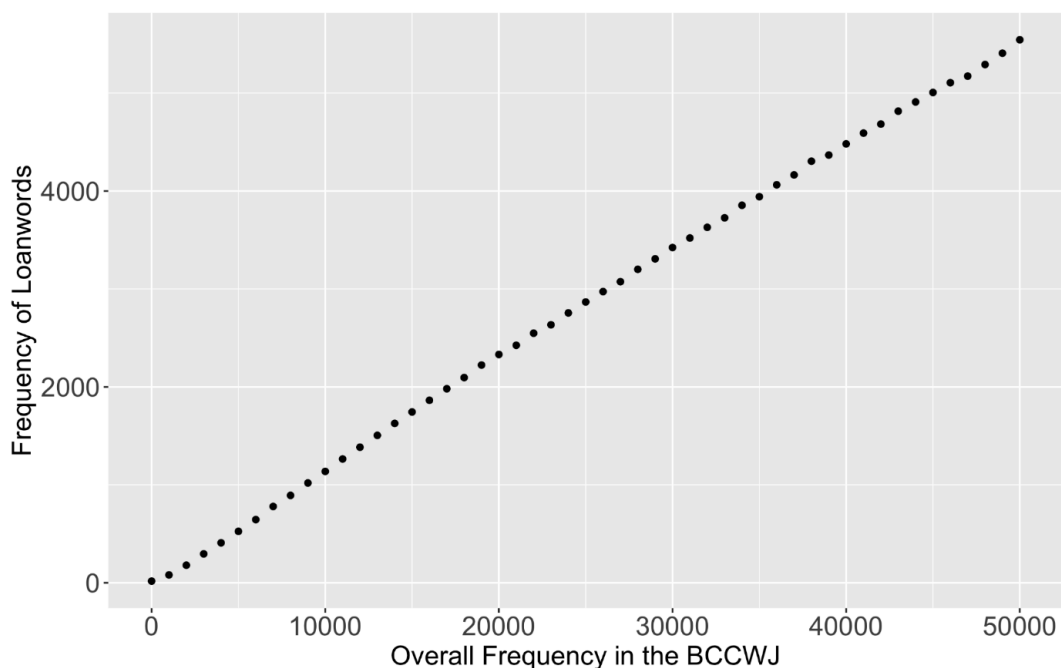


Figure 2.3 Frequency of loanword in every 1000 overall vocabularies in the BCCWJ.



In conclusion, above data have shown that native speakers of Japanese have a vocabulary of around 4,000 loanwords. On the other hand, learners of Japanese are likely to have only around 300 loanwords. If the size of vocabularies plays a significant role in statistical language learning, detecting and tracking likelihood of gemination can be a challenge for L2 learners. In the next section, I outline the research questions and specific predictions for the experiment.

## 2.4 Research Questions and Predictions

Returning to the issues addressed in this thesis, this section states the research questions in more detail. The study focuses on L2 acquisition of phonotactic knowledge throughout the whole thesis. As a reminder, based on previous nativised loanword research, the study considers the following phonological rules of loanwords below before stating research questions

### Rule A: Epenthetic vowels

Because consonants in word-final position are prohibited except with a moraic nasal in Japanese, final consonants in the source language have to be syllabised by epenthesis. The quality of epenthetic vowels is dependent on the preceding phonological context. The sequences [tu] and [du] are illicit in native Japanese. The current study is concerned with only the epenthetic vowels [o] and [u].

- (i)  $\emptyset \rightarrow i / tʃ, dʒ, k \_ \#$
- (ii)  $\emptyset \rightarrow o / t, d \_ \#$
- (iii)  $\emptyset \rightarrow u /$  in all other contexts + k

### Rule B: Consonant Gemination

When monosyllabic English words have the  $C_1VC_2$  phonological structure with a lax vowel in V and a voiceless stop in  $C_2$ ,  $C_2$  will be realised as a geminate consonant.

- (i)  $C_{2[-son, -voice]} \rightarrow CC / V_{[lax]} \_ C_2 \#$

Voiced stops are not supposed to geminate as much as voiceless stops. Alveolar [d] is most likely to undergo geminate than velar [g], and labial [b] is the least likely (i.e., [dd] > [gg] > [bb]) (e.g., Shirai, 1999, Hirayama, 2005, Amano & Kondo, and see §2.3.2.3).

**RQ 1:** Is it possible that a sublexicon phonology of a language is learned from exposure to the target language?

**Prediction 1:**

It is predicted that the learning of phonological rules is possible without being taught. If this is possible, acquisition of phonological rules would differ depending on the degree to which a learner is exposed to Japanese. This is based on the assumption that participants who have more exposure to Japanese have a reasonable level of phonological knowledge due to their accumulated Japanese lexicon.

**RQ2: If any, what rules are implicitly learned?**

**Prediction 2:** For epenthetic vowels (Rule A), there are categorical constraints dictating which vowels should be used. For geminates, within our target sublexicon (i.e., loanwords), there are stochastic patterns, such that the likelihood of different voiced geminates varies. Moreover, voiced geminates are less likely than voiceless geminates in this context. Thus, it is predicted that L2 learners of Japanese are more likely to acquire Rule A than gemination patterns (Rule B) by exploiting their lexicon. This is based on an assumption that categorical rules are more readily learned than gradient ones. However, the epenthetic vowel [u] occurs more frequently in comparison to the other vowel [o] which occurs only after alveolar stops /t, d/. Thus, it would be possible that language users overgeneralise an epenthetic rule in which [u] can be used in any context.

**RQ 3:** Are L2 learners with high exposure to Japanese able to acquire fine-grained knowledge regarding the effects of voicing and place of articulation (POA)?

**Prediction 3:** L2 learners with high exposure to Japanese would be expected to be able to acquire fine-grained knowledge regarding the effects of voicing. The acquisition of this gradient pattern depends on the degree of exposure individuals experience, and the degree of statistical support of the rule. For the effect of POA, [d] is most likely to geminate, while [g] and [b] will have lower ratings for gemination.

**RQ 4:** Does the language's overall statistical patterns influence learners' response patterns? Specifically, would participants be biased in responding with the most expected pattern in the language rather than with observed patterns in the Japanese loanwords?

**Prediction 4:** Participants who do not have access to the knowledge of loanword phonology may access the statistical patterns in the lexicon for their responses, in order to find the most

frequent phonotactic patterns in the language. In that case, singleton would be selected rather than geminates as the overall frequency of geminates are lower in all contexts.

To address the research questions, I conducted an online experiment using a confidence-rating task (i.e., well-formedness task) involving plausible pronunciations of CVC structured English source words. Whether L2 learners are aware of the above rules and patterns observed in nativised loanwords will be investigated in the following chapters.

# Chapter 3

## Pilot Study: Confidence-Rating Task for Adaptation of English Word-Final Consonants

### 3.1 Introduction

The aim of the first experiment is to investigate the extent to which native and non-native Japanese speakers learn Japanese loanword phonology, using the statistical learning mechanism. The adaptation patterns of English word-final consonants in monosyllabic words by epenthesis and gemination would serve as a test case for studying the statistical learning of the sublexicon phonology on well-formedness judgements. More specifically, the current experiment is designed to investigate whether listeners can identify the probabilistic patterns in nativised loanwords.

A confidence-rating task (which is a type of a well-formedness judgment task) was chosen instead of a forced-choice identification task in order to find out whether participants respond more favourably to a specific pronunciation than others. The stimulus word was presented orthographically (e.g., <pip>) to participants as they simultaneously heard one of the pronunciations. Participants were asked to judge whether the pronunciation they heard was how the word would be pronounced if it was a Japanese word, by rating how confident they were on a scale of 1-5. Participants were required to complete the rating task and post-study questionnaire. Details of experiment design are discussed shortly.

In this chapter, I will provide the materials and methodology used in the pilot studies and a brief overview of the outcomes from pilot studies and adaptations to the final research. The main aim of the pilot studies was to validate the research method and to identify issues before the actual study. Eight subjects participated in total. All of them were not university students as participants who are fitted to the current experimental criteria are very limited. Therefore, unfortunately there are different selection criteria for the pilot study. There were three sessions in total. After participants completed the experiment, I received feedback from participants as to whether they found instructions easy to follow, and feedback regarding practical considerations about the experiment. At the same time, the data collected was examined to identify issues with the experiment in order to amend it accordingly. The practical considerations included procedures, wording of instructions, issues related to ratings and the

duration of the session. Whenever amendments were made, two participants tested a new version of the experiment. For example, if instruction wording was changed to avoid confusion for ratings, the revised instruction was tested, and the results examined to evaluate it. The details will be provided shortly and summarised at the end of this section. Since data collected in pilot studies were small, statistical analysis was not conducted for these data.

This chapter has four main sections. Section 3.2 describes the methodology used in the experiment in detail. This contains information about a stimulus speaker and the recordings, and explains how the acoustic analysis was carried out. Section 3.3 provides explanation of pilot tests, their implications and summary of amendments is presented in §3.4.

## 3.2 Research Design and Methodology

### 3.2.1 Materials

Stimuli with monosyllabic English words were selected. The structure of the words is  $[C_1VC_2]$ , V being one of the lax vowels [ɪ, e, æ, ʌ, ɒ]. The consonants in  $C_1$  were any consonant except labiodental and dental fricatives.  $C_2$  will be selected from one of the six stops [p, t, k, b, d, g] (e.g., /pip/, /pek/). Voiced items are included to see whether learners are able to acquire fine-grained knowledge regarding the effects of voicing. Each of the words had five different pronunciations: CVC (*pip*), singleton (CVCV; *pipu*), geminate (CVCCV; *pippu*), wrong epenthesis (*pippo*) and long vowel (CVVCV; *piipo*). Thus, stimulus materials consisted of six sets of 10 CVC English word quintuplets, giving a total of 300 words (Appendix A). The sets were according to the second consonant (i.e., stop consonants) in the words.

The stimuli were classified into three sets for each voicing type, according to the quality of  $C_2$ . For example, in terms of voiceless sets, in the p-Set, a word-final consonant is [p] (e.g., *pip* [pɪp], *nap* [næp], *sup* [sʌp]). In the t-Set, the consonant in coda is [t], like *bet* [bet], *dat* [dæt], and *zit* [zit]. Finally, in the k-Set, words with coda [k] such as *peck* [pek], *tack* [tæk], and *sock* [sɒk] were the target stimuli. Since the purpose of this study is to examine whether participants implicitly learn phonological regularities from English loanwords in Japanese, the words were quintuplets that were fully or partially consistent with loanword phonology. These pronunciation types (i.e., quintuplets) were carefully selected based on previous studies (e.g., (e.g., Preston & Yamagata, 2004; Takagi & Mann, 1994). Examples are shown in Table 3.1. Shading marks the pronunciation type that native speakers of Japanese and L2 learners are expected to give high ratings for the structure corresponding to the source (i.e., target) word, if participants have fully acquired loanword phonological rules. That is, the pronunciation

conforms to all phonological rules in Japanese loanword phonology where each word-final consonant is geminated and yet turns into mora with an appropriate epenthetic vowel. As can be seen in Table 3.1, each target word (i.e., English source word) consists of a quintuplet with the following structure: CVCCV (licit adaption structure), CVCV (phonologically licit but no gemination), CVVCV (phonologically licit but no gemination, and medial vowel modified with vowel lengthening), CVCCV (inappropriate epenthetic vowel), and CVC (no epenthesis and no gemination, but phonemically changed). First and fourth items had an identical structure except for the insertion of an inappropriate vowel for the fourth item. For the /t/-Set, the epenthetic vowel following loanword phonology in this context was the mid-back vowel [o]. In order to create an incorrect modification, the default vowel [u] was added in word-final position, creating the only item which is phonotactically illicit in native Japanese (i.e., \*[tu]). Since inserting [u] in /p/-Set and /k/-Set is phonologically expected, another epenthetic vowel [o] was added for these sets. For voiced items, each target word also consisted of a quintuplet with the following structure: CVCV (licit adaption structure), CVCCV (phonologically licit but with gemination), CVCV (inappropriate epenthetic vowel), CVVCV (phonologically licit but no gemination, and medial vowel modified with vowel lengthening), and CVC (no epenthesis and no gemination, but phonemical changed). For the voiced items shown in Table 3.2, shading marks the pronunciation type that L2 learners are expected to give high ratings for the structure that differs from the voiceless items. This is because voiced stops are not supposed to geminate. For the voiced context, the pronunciation conforms to all phonological rules in Japanese loanword phonology where each word-final consonant is singleton and yet turns into mora with an appropriate epenthetic vowel.

Note that to select English words as stimuli, the BCCWJ (see §2.3.2.3 for more details) was used to select English words that have not yet been nativised in Japanese or nativised words with low frequency. The loanword with the highest frequency was 605.97 words per million (i.e., a loanword *paasento* from the English ‘percent’: 633,992 (frequency) /104,612,423 (total frequency in the corpus) \*1 million=605.97). Lower frequency is defined as 0.02 words per million in this paper. Those selected words are also of low frequency in CELEX (Baayen, Piepenbrock, & Van Rijn, 1993).

Table 3.1 Examples for sets of stimuli and their structures: voiceless sets

Word Structure		Original Structure	Modification Types: Representation of the target word						
		CVC	CVCCV	CVCV	CVCCV	CVCCV		CVC	
		1 syllable	3 morae	2 morae	3 morae	3 morae		1 syllable	
Target Word (Source Word)		conforming to loanword phonology	no gemination	vowel lengthening	inappropriate epenthetic vowel		only vowel phoneme substitution		
					unexpected vowel	default vowel [u]			
Phonological Regulation in Loanword	Gemination	N/A	✓	✗	✗	✓	✓	✗	
	Quality of Epenthetic Vowel	N/A	✓	✓	✓	✗	✗	✗	
	Phonologically Licit	N/A	✓	✓	✓	✓	✗	✗	
Lax Vowel	Key Word								
			<b>/p/-Set</b>						
[ɪ]	KIT	pɪp[pɪp]	/pippu/	/pipu/	/piipu/	/pippo/	N/A	/pip/	
[æ]	TRAP	dap[dæp]	/dappu/	/dapu/	/daapu/	/dappo/	N/A	/dap/	
			<b>/t/-Set</b>						
[e]	DRESS	bet[bet]	/betto/	/beto/	/beeto/		/bettu/	/bet/	
[ʌ]	STRUT	jut[dʒʌt]	/dʒatto/	/dʒato/	/dʒaato/		/dʒattu/	/dʒat/	
			<b>/k/-Set</b>						
[ʌ]	STRUT	puck[pʌk]	/pakku/	/paku/	/paaku/	/pakko/	N/A	/pak/	
[ɒ]	LOT	bock[bɒk]	/bokku/	/boku/	/booku/	/bokko/	N/A	/bok/	

Table 3.2 Examples for sets of stimuli and their structures: voiced sets

		Original Structure		Modification Types: Representation of the target word				
		CVC	CVCCV	CVCV	CVVCV	CVCV	CVC	
Word Structure		1 syllable	3 morae	2 morae	3 morae	2 morae		1 syllable
		Target Word (Source Word)	gemination	conforming to loanword phonology	vowel lengthening	inappropriate epenthetic vowel		only vowel phoneme substitution
						unexpected vowel	default vowel [u]	
Phonological Regulation in Loanword	No Gemination	N/A	✗	✓	✗	✓	✓	✗
	Quality of Epenthetic Vowel	N/A	✓	✓	✓	✗	✗	✗
	Phonologically Licit	N/A	✓	✓	✓	✓	✗	✗
Lax Vowel	Key Word							
		<b>/b/-Set</b>						
[ɪ]	KIT	nib[nɪb]	/nibbu/	/nibu/	/niibu/	/nibo/	N/A	/nib/
[æ]	TRAP	nab[tæb]	/nabbu/	/nabu/	/naabu/	/nabo/	N/A	/tab/
		<b>/d/-Set</b>						
[e]	DRESS	med[med]	/meddo/	/medo/	/meedo/	N/A	/medu/	/med/
[æ]	TRAP	tad[tæd]	/taddo/	/tado/	/taado/	N/A	/tadu/	/tad/
		<b>/g/-Set</b>						
[e]	DRESS	keg[keg]	/keggu/	/kegu/	/keegu/	/kego/	N/A	/keg/
[ɒ]	LOT	bog[bɒg]	/boggu/	/bogu/	/boogu/	/bogo/	N/A	/bog/



### 3.2.2 Audio stimuli

The audio stimuli for a perception experiment were first created by recording a female native speaker of Japanese who is a colleague at the Linguistics Department and has studied linguistics to doctoral level. She speaks standard Japanese (a variety used in educational settings and public broadcasting). The recording took place in a sound-attenuated room at the University of Canterbury, using a Tascam HD-P2 audio recorder with 44,100 samples/s, 16 bit/s and a Beyerdynamic head-mounted microphone. The stimulus speaker's participation was approved by the University of Canterbury Human Ethics Committee (under application number HEC 2018/29/LR-PS).

For the recording, the stimuli were produced in the carrier sentence in Japanese characters, *Korewa \_\_\_\_\_ desu.* 'This is \_\_\_\_.' PowerPoint slides were used for displaying stimuli, with one slide for each sentence. Each target word was displayed in *katakana* but other words are in *hiragana*. This is because *katakana* is conventionally used for loanwords. The purpose of using the carrier sentence was to maintain the same tempo, intensity and tone across readings. A tone pattern of all target words was a HL (high-low) sequence. This is default accent for loanwords in Japanese such as the English words 'kick', 'pot', 'bed' in Japanese as /'kikku/, /'potto/, and /'beddo/ (Kindaichi & Akinaga, 2014), and is commonly used for singleton-geminate discrimination experiments (e.g., Asano, 2018; Hardison & Saigo, 2010). The speaker was asked to say as naturally as possible each stimulus including the carrier sentence after it was presented on a computer screen. She repeated each sentence three times. Then, production recordings were analysed acoustically using Praat phonetic software (Boersma & Weenink, 2018; hereafter Praat). For each stimulus token, the best recording was manually selected from the three options, giving consideration to clarity of production and the duration of words. Finally, the target words were extracted from the carrier sentence. To validate the stimuli, 10 words were selected and created into two concatenated sound audio clips that were listened to by four colleagues in the Linguistics Department at the University of Canterbury. They agreed that phonological features such as singleton vs. gemination and long vowel vs. short vowel were clearly present, however, some vowels in word-final position were perceived as being cut off. Therefore, another recording was conducted.

For the second recording, each word was produced without a carrier sentence. Each PowerPoint slide was advanced automatically after three seconds to help keep the same tempo for the reading of each word. For each stimulus, the best recording was manually selected from the three options. In order to compare the quality of stimuli between the previous recording

and the second recording, exactly the same 10 words as the first recording were selected and concatenated into an audio clip. It was assessed by the same four people so that the second recording did not contain any problems like the previous recording. Therefore, this recording was used instead of the first recording and the selected sounds were extracted from the recording files.

Eleven stimuli of the form CVC had higher intensity than the other stimuli, therefore, the intensity of other CVC tokens was measured, and their mean calculated. Then, the intensity of the eleven stimuli was modified to ensure consistency across stimuli. There were a total of 150 audio files for each voicing type.

### 3.2.3 Acoustic characteristics of the audio stimuli

In order to determine the acoustic characteristics of the stimuli, all vowels in the audio stimuli were analysed using Praat (Boersma & Weenink, 2018). The duration of each target vowel was measured and the mean values for F1, F2, and F3 for the stimulus vowels were extracted using a custom Praat script. All measurements were taken at the midpoint of the marked segment. All extracted formants were checked manually to ensure the validity of the values. Note that /i/ and /u/ are not devoiced in the environments of preceding voiceless consonants or between two voiceless consonants.

First, vowel formants are considered. The mean values for F1, F2 and F3 for the stimulus vowels and their standard deviation (SD) are shown in Table 3.3.

Table 3.3 Mean F1/F2/F3 formant values and standard deviations for voiced and voiceless contexts

Number of Token	Vowel	Symbol for Plot	Position in Word	F1		F2		F3	
				mean	SD	mean	SD	mean	SD
88	[a]	a	medial	887.1	40.7	1657.3	101.6	2830.2	222.2
22	[aa]	aa	medial	907.3	34.5	1555.3	52.6	3111.9	145
36	[e]	e	medial	527.6	60.4	2366.8	88	3138.7	71.1
9	[ee]	ee	medial	563.6	56.1	2366.2	55	3176.1	54.4
40	[i]	i	medial	291.7	22	2727.4	83	3552.6	179.1
10	[ii]	ii	medial	277.6	13.5	2795	43.9	3684.4	188.3
76	[o]	o	medial	490.6	47.1	988	125.2	3150.0	196.8
100	[o]	O	final	441.9	46	856.8	56.6	3009.6	180.3
19	[oo]	oo	medial	459.6	47.1	844.6	79.1	3201.8	150.2
140	[u]	u	final	360	46.2	1428.4	90.9	2939.1	107.1

Figure 3.1 shows the individual vowels from the speaker. Since only one speaker was used to create stimuli, the data was not normalised. For this context, the vowel /o/ in word-final position is indicated as the capital O. As can be seen, vowels in the plots do not overlap with each of the other five vowel categories. The F1/F2 space for the stimulus vowels is consistent with the Japanese vowel space presented in Vance (2008). The high front vowel [i] is higher and fronter than other vowels. Although the other high vowel [u] is slightly centralised, this is compatible to the findings by Nogita, Yamane, and Bird (2013). The mid-front vowel [e] and mid-back vowel [o] are similar in terms of height. The vowel [a] is centralised in backness.



Figure 3.1 F1 and F2 values for individual vowels from the stimulus speaker.

Secondly, vowel duration is considered. Table 3.4 shows descriptive statistics for vowels produced by the speaker. The standard deviation of vowel durations for long vowels ranges from 39 - 49 ms, and for short vowels ranges from 18-25 ms, consistent across vowels. The ranking of vowels from longest to shortest is [aa] [oo] [ee] [ii] followed by [a], [o], [e], [i] in word-medial position, and [o] [u] in the word-final position. The results are fairly consistent with earlier vowel duration studies (Campbell, 1992; Han, 1962 cited in Shoji & Shoji, 2014). Although [u] is acknowledged as the shortest vowel in Japanese, cross-linguistically vowels in

utterance final position are normally longer than vowel in non-final positions (Johnson & Martin, 2001), and final mora is lengthened in Japanese (Kaiki, Takeda, & Sagisaka, 1990).

Table 3.4 The number of token counts, the mean and SD in ms. of the five vowels along with vowel-to-word duration ratio

Number of Tokens	Vowel	Symbol for Plot	Position in Word	Vowel Duration (ms)		Vowel-to-Word Duration Ratio
				Mean	SD	Mean
88	[a]	a	medial	130	25	0.237
22	[aa]	A	medial	342	49	0.486
36	[e]	e	medial	127	22	0.231
9	[ee]	E	medial	332	44	0.484
40	[i]	i	medial	106	18	0.187
10	[ii]	I	medial	311	48	0.447
76	[o]	o	medial	128	18	0.231
100	[o]	O	final	174	19	0.287
19	[oo]	oo	medial	341	39	0.490
140	[u]	u	final	173	20	0.275

Figure 3.2 shows differences in duration across vowel qualities. White circles indicate mean values in the box plots. As can be seen in the plot, the box plots do not overlap between long and short vowels for each vowel quality, which indicates that they tend to be significantly different with a 95% confidence level (McGill, Tukey, & Larsen, 1978). Since the figure clearly shows the differences between long vowels and short vowels, vowel-to-word duration ratios were not statistically analysed.

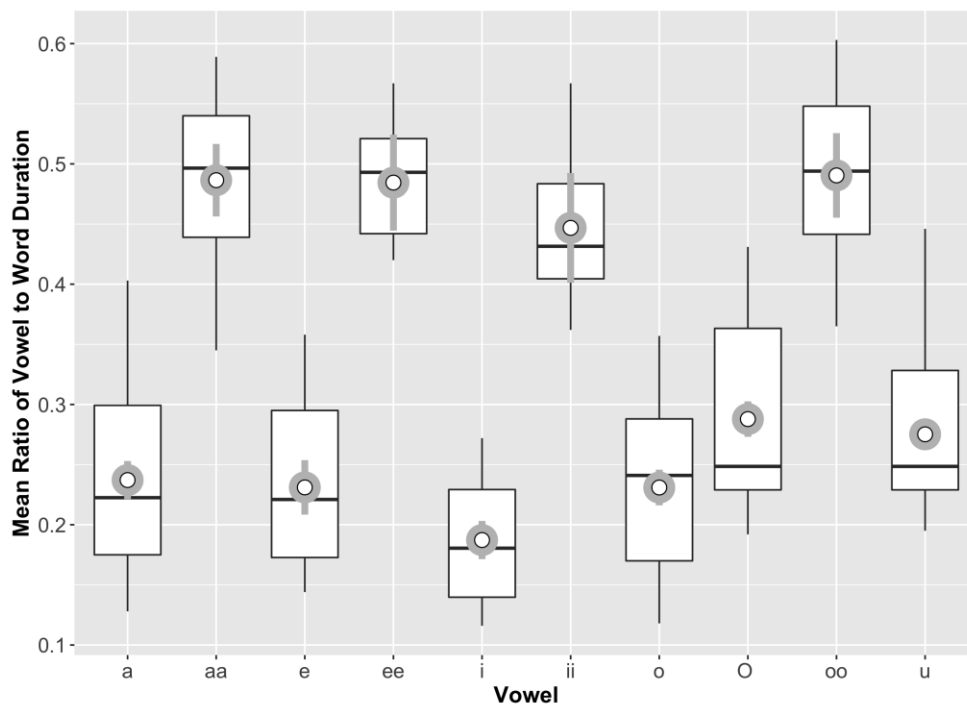


Figure 3.2 Boxplots of the ratio of vowel-to-word durations for the stimulus speaker.

Figure 3.3 shows differences in duration across closure qualities between singleton and gemination words, for voiceless in the top panels and for voiced stimuli in the bottom panels. Note that for the measurement, VOT is excluded. Again, as can be seen in the plot, the box plots do not overlap between geminated and singleton consonants, which indicates that they are significantly different in duration.

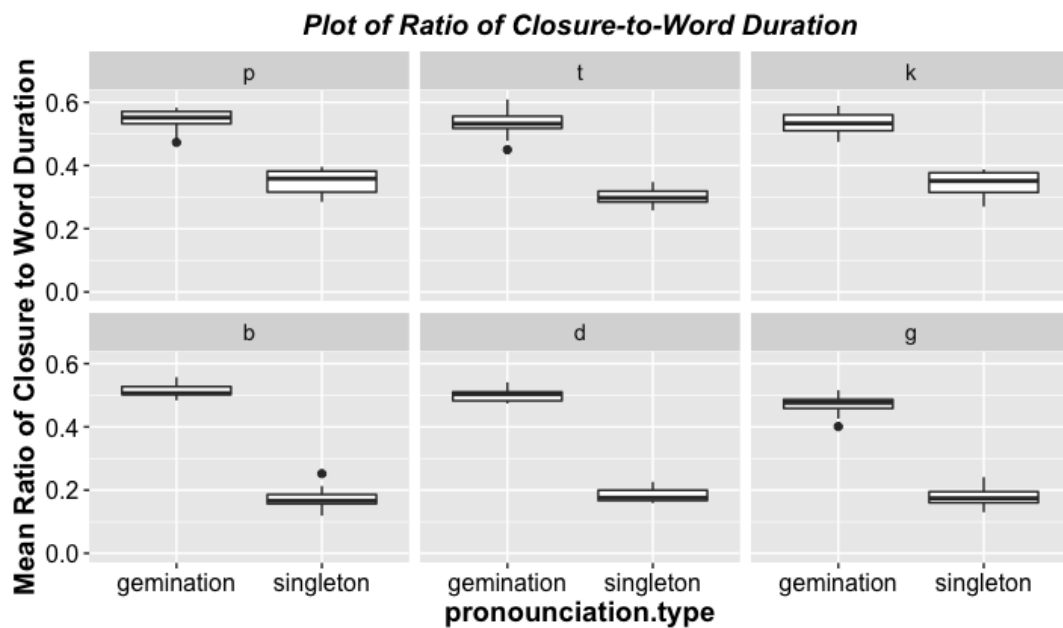


Figure 3.3 Boxplots of the ratio of closure-to-word duration for the stimulus speaker.

According to Hirata (2007), one of the best index parameters for reliable stop quantity distinction is the duration ratio of gemination and singleton closure. Hirata and Whiton (2005) found that the durational ratio of gemination and singleton closure was approximately 3:1 for voiceless gemination. For the current study, the duration ratio of closure duration between gemination and singleton stops varies across place of articulation and voicing contexts. For voiceless, [p]/[pp] is 2.53, [t]/[tt] is 2.77, and [k]/[kk] is 2.48, which is compatible with previous studies on closure duration (Han, 1994; Hirata & Whiton, 2005; Homma, 1981). For voiced context, the ratios for places of articulation are longer than in the voiceless context; [b]/[bb] is 4.83, [d]/[dd] is 4.9, and [g]/[gg] is 4.26. The ratios in the voiced context are higher than in the voiceless context, which is consistent with the findings in previous studies on closure durations of singleton and geminate stops (e.g., Hirose & Ashby, 2007; Homma, 1981; Kawahara, 2006). Important here is that there is clear difference between geminate and singleton stimuli.

### **3.2.4 Participants**

A total of eight participants (two native New Zealand English (NZE)-speaking learners of Japanese, and six native speakers of Japanese) completed the pilot studies and the three versions of the speech perception experiment in total. Firstly, two NZE-speaking learners of Japanese (one female and one male), 18-24 years of age, and two female native speakers of Japanese aged over 40 years participated. One of the learners had recently completed a three-year Japanese course at the University of Canterbury and the other learner was studying Japanese at a high school at the time data was collected. For the following two sessions, two female native speakers of Japanese aged over 40 years participated for each session. As for all of the Japanese listeners, their total years living in New Zealand (NZ) was more than five years. The recruitment procedures and all the text used were approved by the University of Canterbury Human Ethics (under application number HEC2018/29/LR-PS).

### **3.2.5 Procedure**

The experiment was implemented as an online rating task using an online survey platform called the Speech In Noise 2 (Chan, 2018) via the NZILBB (New Zealand Institute of Language, Brain and Behaviour) link. Each participant did the task in their own convenient time and place. The entire procedure was described to participants in English. For Japanese speakers, instructions were written in both English and Japanese. However, the experimental page was written only English.

At first, participants completed the audio system test in which they were instructed to listen to the stimuli through headsets. Two audio files, ‘dog’ and ‘book’, created by a NZ male speaker were presented. After participants listened to the audio files one at a time, they typed the word they heard. Then, following instructions on the computer screen, they were asked to complete the confidence rating task. All the pronunciations from 60 quintuplets as stimuli were presented in a different random order for each participant, for a total of 300 trials (150 voiceless, 150 voiced). In each trial, the stimulus word was presented orthographically (e.g., <pip>) on the computer screen to participants as they heard one of the pronunciations (e.g., /pipu/, /pippu/, /piipu/, /pippo/, /pip/) through headsets. After hearing the stimulus (e.g., /pippu/), participants were asked to judge whether the pronunciation they heard is correct pronunciation in Japanese if this was a Japanese word, by rating how confident they are on a scale of 1-5. The scale is defined as follows: 1= “confident that this is NOT a correct pronunciation in Japanese”, 5= “confident that this IS a correct pronunciation in Japanese”. After the participant clicks one of

the options and a 'NEXT' button, the next stimulus was presented. Audio stimulus would be played only once. After finishing the rating task, the participants had to fill out a questionnaire (described in §4.1.4). There was no training phase or practice phase. No feedback was given during the experiment.

### **3.3 Overall Outcome and Application to Actual Experiment**

#### **3.3.1 Duration of the task**

The duration for completing the task (including responding to the questionnaire) was estimated at approximately 40 minutes in total. During the initial pilot study, Japanese participants completed the task in less than 40 minutes, whereas one of NZE-speaking learners of Japanese took approximately 50 minutes and the other learner took more than one hour. Since the questionnaire for learners of Japanese was longer than the questionnaire for native speakers of Japanese, it was expected that learners would take a longer time than the other group. As the task was quite simple, participants might have found it tedious to finish and could have been distracted by other things. Therefore, I considered making the task shorter to allow respondents to finish faster. In order to shorten the task, the pronunciation with long-vowel (i.e., 60 CVVCV stimulus) were excluded. The long-vowel stimuli were present to allow English-speaking learners to choose this option instead of geminates (i.e., CVCCV) if they have learned Japanese prosodic patterns (e.g., Preston & Yamagata, 2004). However, since this is not a main question category in this perception study, this variety was excluded from the next session. In the actual real test, the confidence rating task would contain 60 quadruplets for a total of 240 trials, instead of 60 quintuplets for a total of 300 trials. As a result, subsequent tests were completed in less than 30 minutes.

#### **3.3.2 Rating**

##### **3.3.2.1 CVC tokens**

For the first pilot test, the results show that while NZE-speaking learners of Japanese rated CVC stimuli low, native speakers of Japanese rated these stimuli higher than the learners. Because NZE-speaking learners rated CVC stimuli low, CVC tokens were explicitly evaluated as CVC. This posed a concern about whether Japanese participants misunderstood the experiment as seeking to access their English ability rather than their Japanese phonological knowledge. This ratings issue might be attributed to couple of things. Firstly, for native speakers of Japanese, introductions were written in both English and Japanese. Afterwards, all

instructions used only Japanese. Secondly, prior to the experimental phase, an audio system test took place during which participants listened to two audio clips, one at a time, and they were asked to type what they heard (as described in procedure (§3.2.5)). This phase is a simple audio test, however, after typing the word they heard, a pop-up screen “correct” appeared. It is not preferable to make this influence participants. Therefore, the audio clips by an English speaker were replaced with that of a Japanese speaker, saying ‘panda’ and ‘sushi’, in order to avoid the impression this is an English listening test for Japanese listeners.

After the amendment, four participants tried the new version. It was observed that three participants rated CVC low almost all the time. However, one participant still rated the CVC stimuli ‘4’ (out of five) consistently. Therefore, the rating instruction with examples was added for the future experiment as follows. This instruction was given in Japanese for native Japanese speakers, and in English for native English speakers.

In the case of 'MacDonald' second-language learners of Japanese might say one of these (a) 'makudonarudo', (b) 'makudonadu', or (c) 'macdonald'. Neither (b) or (c) would be the correct way to pronounce this in Japanese. If you know this, then you would give both a rating of '1', and you would give the correct pronunciation (a) a rating of '5'. Note that it is not relevant that 'Makudonadu' is slightly closer than 'Macdonald' - if you are equally confident that both are wrong, you should give them both the same rating.

This rating issue for CVC tokens by Japanese listeners could be due to the phonotactics of Japanese. As discussed in Chapter 2, it is well known that native speakers of Japanese tend to perceive illusory vowels inside consonant cluster (e.g., *ebzo*) in stimuli, even when no vowel was presented (Dupoux et al., 1999; Dupoux, Parlato, Frota, Hirose, & Peperkamp, 2011). An epenthesis effect was observed regardless of linguistic background of the listeners (Dupoux et al., 1999). The results revealed no clear effect of self-reported levels of proficiency in English or French which allows consonant clusters. Research indicates that this is due to the phonology of listeners’ native language which constrains the perception of non-native syllable structures. In a follow-up study by Monahan, Takahashi, Nakao, and Idsardi (2009), Japanese listeners also have difficulty discriminating between [eguma]/[egma] and [ekuma]/[ekma]. It is not known whether the same effect can be observed at the word-final position. However, if listeners’ native language phonology constrains the perception of consonant clusters, the same effect would be expected on the word-final consonant as well. The reason is that both of them are



illicit syllable structures in Japanese. The Japanese listeners in this pilot study were not naïve to English, as they have been living in NZ for more than five years. However, the length of residence does not guarantee their listening ability for the non-native sounds. This rating issue needs to be kept in mind.

### **3.3.2.2 Interpretation of “correct”**

Feedback from one participant was very useful, in that the phrase ‘a correct pronunciation’ could guide people towards focussing on overall phonotactics, rather than the phonotactics for the specific loanword adaptation. For example, a word ‘nip’ is supposed to be pronounced as ‘nippu’ following Japanese loanword phonology, therefore ‘nippo’ is not ‘a correct pronunciation’. However, /nippo/ is not only phonetically correct but also frequently occurs in Japanese, such as ‘nippon’ which means ‘Japan’. Therefore, wording in the choices were amended as follows:

- 5 Confident that this IS how the word would be pronounced in Japanese
- 4 Somewhat confident that this IS how the word would be pronounced in Japanese
- 3 I really do not know if the word would be pronounced this way or not
- 2 Somewhat confident that this is NOT how the word would be pronounced in Japanese
- 1 Confident that this is NOT how the word would be pronounced in Japanese

## **3.4 Summary**

The aim of the pilot study was to validate the research method and design of the questionnaire, and to identify issues before the actual study. The aspects that changed as a result of the pilot study are summarised in Table 3.5. Some of the major changes are (1) removing 60 CVVCV tokens from 300 stimuli, making the duration of experiment shorter; (2) changing from five pronunciations to four; (3) adding an explanation for the ratings procedure, with examples provided; (4) changing the description for Japanese listeners from both English and Japanese to only Japanese; (5) changing the wording for pronunciation well-formedness in order to avoid confusion between overall phonotactics and loanword phonotactics; (6) changing speakers for audio clips for the audio system test.

Table 3.5 Summary of amendment after pilot study

Area of Change	Change Effected
Number of stimuli	<p>300 →240</p> <p>Removing 60 tokens of the CVVVCV pronunciation, the total trials would be 240 from an initial 300. This reduces the duration of the experiment, which would be expected to be completed within 30 minutes.</p>
Pronunciation type	<p>quintuplets → quadruplets (CVC, CVCV, CVCCV, *CVCCV)</p> <p>As a result of removing the CVVVCV tokens, there would be four pronunciations for each target word.</p>
Explanation of rating	Explanation for rating with examples is added.
Description for Japanese listeners	Written in both English and Japanese → in only Japanese
Interpretation of well-formedness	<p>Wording was amended.</p> <p>‘this is a correct pronunciation in Japanese’ →‘would be pronounced in Japanese’</p>
Speaker for audio files of audio system test	<p>English speaker → Japanese speaker</p> <p>Replacing audio files by an English speaker with that of a Japanese speaker, in order to avoid the impression this is an English listening test for Japanese listeners.</p>

## Chapter 4

# Confidence-Rating Task: Adaptation of English Word-Final Consonants

The findings of the pilot studies described in the previous chapter were considered and some amendments on methodology were made for the actual experimental study which is presented in this chapter. This experiment investigates whether listeners can identify the probabilistic patterns in nativised loanwords. A fully-crossed online investigation into adaptation of English final consonants was conducted with 22 NZE-speaking learners of Japanese, 20 NZE-speaking non-learners of Japanese and 20 Japanese listeners.

As reminder, this study asks following questions: **(Q1)** is it possible that a sublexicon phonology of a language is learned from exposure to the target language? **(Q2)** If any, what rules are implicitly learned in relation to English word-final stop consonants? (i.e., epenthetic rule or geminates or both) **(Q3)** Are L2 learners with high exposure to Japanese able to acquire fine-grained knowledge regarding the effects of voicing and place of articulation (POA)? and **(Q4)** Does the language's overall statistical patterns influence learners' response patterns? Specifically, would participants be biased in responding with the most expected pattern in the language rather than with patterns they observed in the Japanese loanwords?

This chapter has six main sections. Section 4.1 describes the methodology used in the experiment in detail. It includes information about phonotactic scores of overall Japanese and that of loanwords in experimental stimuli were calculated based a dictionary corpus created from the data of the Corpus of Spontaneous Japanese (henceforth CSJ, National Institute for Japanese Language and Linguistics, 2017). This section also presents Principal Component Analysis (PCA) for questionnaires. Section 4.2 presents statistical analysis and results phonotactic effects extracted from the CSJ. Then, factors and statistical models for phonological process are presented in §4.3. Results of the experiment analysis, indicating how and to what extent the groups of participants differ in their response patterns for each pronunciation type are provided in §4.4, followed by discussion and a subsequent research direction is offered in §4.5. Finally, a conclusion is presented in §4.6.

## 4.1 Research Design and Methodology

### 4.1.1 Participants

University students in NZ with no history of a speech or hearing disorder participated in the task. There were three groups of listeners, a total of 65 students: (1) native speakers of Japanese, (2) native speakers of NZE taking Japanese as L2 courses at university, and (3) native NZE-speaking non-learners of Japanese. The group of native Japanese speakers serves as a baseline group. For the English speakers, participants in the experiment were native speakers of NZE. Both learners and non-learners of Japanese are chosen to examine how well people understand the pronunciation characteristics and sound rules in a language different to their native language. NZE-speaking learners of Japanese have knowledge of Japanese as L2 learners. Non-learners are NZE-speaking monolinguals who are chosen because they do not have formal knowledge of Japanese language. All of the participants were tested on the same stimuli. Thus, by comparing learners' group with non-learners' group, I can investigate how much knowledge of Japanese sound rules L2 learners of Japanese actually have.

All of the participants were recruited via social-media networks, word-of-mouth, and through on-campus recruiting at the end of the academic year. For the learners of Japanese, participants were mainly recruited before/after their Japanese classes by receiving recruitment flyers. Participants were informed that the experiment would take approximately 30 minutes, for which they received a payment of NZD 10 e-voucher. The recruitment procedures and materials used were approved by the University of Canterbury Human Ethics Committee (under application number HEC2018/29/LR-PS).

Two participants were excluded for not matching their native language to the selection criteria, and two participants were excluded as they checked the box indicating speech or hearing impairment, leaving a total of 62 participants whose responses were analysed. The number of participants in each group is presented in Table 4.1. Although ideally, learners of Japanese would have been more, there was not a large enough body of NZE speakers enrolled in Japanese courses to draw from.

Table 4.1 Number of participants in each group for analysis

<b>Participants</b>	<b>Number</b>
Native Speakers of Japanese	20
NZE-Speaking Learners of Japanese	22
NZE-Speaking Non-Learners of Japanese	20
Total	62

## 4.1.2 Materials

### 4.1.2.1 Stimuli

The same material as in the pilot study was used except the long vowel stimuli which were excluded from the stimuli as a result of the pilot study. There were 60 CVC structured English words and each of them has four different pronunciations: CVC (*pip*), singleton (CVCV; *pipu*), geminate (CVCCV; *pippu*) and wrong epenthesis (*pippo*). Thus, there were 60 quadruplets for a total of 240 trials; 120 audio stimuli for each voicing type, as laid out in Table 4.2. A full list of perceptual stimuli is given in Appendix B.

Table 4.2 Number of audio stimuli for each participant

Voicing Type	C2	Pronunciation				Number of Trials
		CVC	Geminate	Singleton	Wrong epenthesis	
voiceless	[p]	10	10	10	10	120
	[t]	10	10	10	10	
	[k]	10	10	10	10	
voiced	[b]	10	10	10	10	120
	[d]	10	10	10	10	
	[g]	10	10	10	10	
Total		60	60	60	60	240

### 4.1.2.2 Extracting for statistical phonotactic scores of Japanese

Next, the investigation of the properties of a spoken language corpus from a large sample can further reveal the response patterns in the current experiment. The Corpus of Spontaneous Japanese (henceforth CSJ, National Institute for Japanese Language and Linguistics, 2017; Maekawa, Koiso, Furui, & Isahara, 2000) contains spoken language, which enabled me to access annotated phonemic information directly. Phonotactic probabilities were calculated over triphones of phonemes of words in the data, by which phonotactic scores for each stimulus item in the present study were created. This informed us how phonotactically similar and different loanwords in Japanese and overall Japanese words are.

The CSJ is a large annotated speech corpus which contains more than 650 hours spontaneous standard Japanese by 1,418 speakers, therefore useful for establishing baseline data on the Japanese language in general. As noted above, loanword phonology is a sublexicon phonology of Japanese phonology. Similarity and frequency effects based on the corpus might

offer some important insights into relationship between loanword phonology and native phonology (Crawford, 2009). Before extracting data for the current study, taking into consideration previous research which used the CSJ, Okada (2008) reported the ratio of V or CV morae (e.g., /a/, /ki/, /te/) to that of special morae (including the moraic nasals (i.e., /N/), the geminate consonants, and the long vowels) in Japanese as 1: 0.198 (cited in Takeyasu, 2010). Takeyasu (2010) undertook further analysis of the special morae in Okada's data. He reported the proportion of each category as being 15% geminate consonants, 31% moraic nasals, and 54% being long vowels. Thus, overall, occurrence ratio of geminates is very low in Japanese.

The current study used the CSJ core dataset known as the CSJ-Core, which contains phonemic and phonetic information in the morphological unit. The purpose of using the corpus is to create a lexical database of Japanese, in order to calculate statistical probabilities of stimulus words in an experiment. To do so, the transcriptions of the speech by 100 speakers were used as a subset. The majority of the speech data are in the Academic Presentation Style by 70 speakers along with Stimulated Public Speaking by 18 speakers and dialogue by 12 speakers. Using the information provided by the CSJ, words tagged as fillers and speech errors (i.e., wrong pronunciations) were removed at an initial stage of the data screening. A total of 244,239 words were extracted based on SUW (short-unit word), approximating the entry-form of Japanese dictionary (Maekawa, Koiso, Furui, & Isahara, 2000). In accordance with the aim of the current analysis, type frequency is of relevance rather than token frequency, therefore, only unique words were extracted using R (R Core Team, 2018). By doing this process, duplicate words such as homophones (e.g., /kata/ (HL) 'shoulder' and /kata/ (LH) 'shape') were removed.

In the original corpus, phonetically palatalised sounds 'gj' (i.e., /g/ before /i/) are assigned to different categories from 'g' (i.e., /g/ before elsewhere). However, they were treated as the same phonemic category as in the current data (i.e., /g/). On the other hand, phonologically palatalised consonants such as 'gy' (e.g., /gjo/ 'fish' vs. /go/ 'word') were treated as a different phonemic category.

All words take dictionary-cited forms in the current dataset. That is, all Japanese verbs and adjectives are base forms, and all inflectional forms were removed or replaced by base forms. All words were further coded as loanword, abbreviation/symbol, blend and others (i.e., either native Japanese, Sino-Japanese, or mimetics). Words coded as symbols and abbreviations such as 'OHP', and 'IPA' were excluded from the sets. Ambiguous words were always checked against the BCCWJ, and I followed its word type categories. Pronouns are

included in data only when the words occur in dictionaries. Blend words between Japanese and other languages are decomposed. For example, in 輪ゴム /wagomu/ ‘rubber band’: while /gomu/ ‘gum/rubber’ is a loan word, /wa/ is a native Japanese word. Therefore, /gomu/ was coded as a loanword and /wa/ was coded as others. As a result, 6,932 unique words remained as a dictionary dataset, of which 1,108 were loanwords. Finally, English words used metalinguistically are not included in loanwords.

Next, these words were coded appropriately to calculate tri-phones. For example, all of the first parts of geminate consonants were transcribed in a conventional manner as a mora phoneme “Q” (e.g., /iQpai/ ‘many’, /moQto/ ‘more’), instead of /i<sup>̄</sup>ppai/ and /mo<sup>̄</sup>tto/. This can create ambiguity as to what phoneme occurs in “Q”. Therefore, all existing geminates were coded as one character to distinguish from singleton consonants (e.g., pp=1, tt=2, bb=0). For example, /iQpai/ was replaced with /i1ai/ and /motto/ was replaced with /mo2o/. This distinguishes the place of articulation in ‘Q’ as well as indicating a simple /p/ from the first part of the geminate /pp/. Moriac nasal was kept as /N/ regardless of its following consonants since nasals are not the focus of the current study.

In order to calculate statistical phonotactics and smoothing scores, a SRILM tool kit (Stolcke, 2002) was used. It is an open-source toolkit for training n-grams and applying statistical language models. Since /bb/ (i.e., 0) did not appear in the dictionary dataset, a loanword スノッブ/sunobbu/ ‘snob’ is added to avoid log 0. This word is in the BCCWB as well as another loanword dictionary (Nobu, Vardaman, & Imidas, 2006). In order to extract dictionary-based phonotactic scores, phonemic representations of 6,932 words were used to generate tri-phone scores. Based on the tri-phone score, the probability of the stimulus words (240 words) in the current experiment was log transformed with base 10. Since scores are not length-normalised, the shorter words should tend to have better scores.

Based on the dataset, the extracted phonotactic scores were applied to experimental stimulus and used for statistical analyses. This addresses a question as to whether statistical patterns in the lexicon influence the response patterns for loanword adaptation over phonological rules. This alternative interpretation will be investigated by comparing the response patterns in the following experiment to phonotactic scores in the current data based from the CSJ data.

#### 4.1.2.3 Phonotactic probability of stimuli

In order to assess the relationship between the loanword phonotactic pattern in Japanese and overall Japanese phonotactic pattern (including Sino-Japanese and loanwords), a Pearson product-moment correlation coefficient was computed using the phonotactic scores of all stimulus pronunciation, based on the data from the CSJ. As a reminder, in order to extract dictionary-based phonotactics scores, phonemic representations of 6,932 words were used to generate tri-phone scores. Scores are not length-normalised, therefore the shorter items should tend to have better scores. As mentioned before, a triphone phonotactic score of all stimulus items (i.e., 240 items) was log transformed. Figure 4.1 illustrates the correlation between the log-phonotactic scores for overall Japanese and that of loanwords under two voicing contexts. The x-axes ‘loanLog’ stand for phonotactic scores of loanwords in Japanese, whereas the y-axes ‘overallLog’ stand for that of overall Japanese phonotactics. The left panel shows the correlation in the voiceless context and the right panel shows that of the voiced context. The plots repel overlapping text labels for readability. Pearson’s correlation tests reveal a statistically significant correlation between the two variables in the voiceless context ( $r = 0.66$ ,  $p < .001$ ) and voiced context ( $r = 0.88$ ,  $p < .001$ ). In these contexts, increases in loanword phonotactics scores were correlated with increases in higher overall phonotactic scores. The tendency is stronger in the voiced type than in the voiceless context.



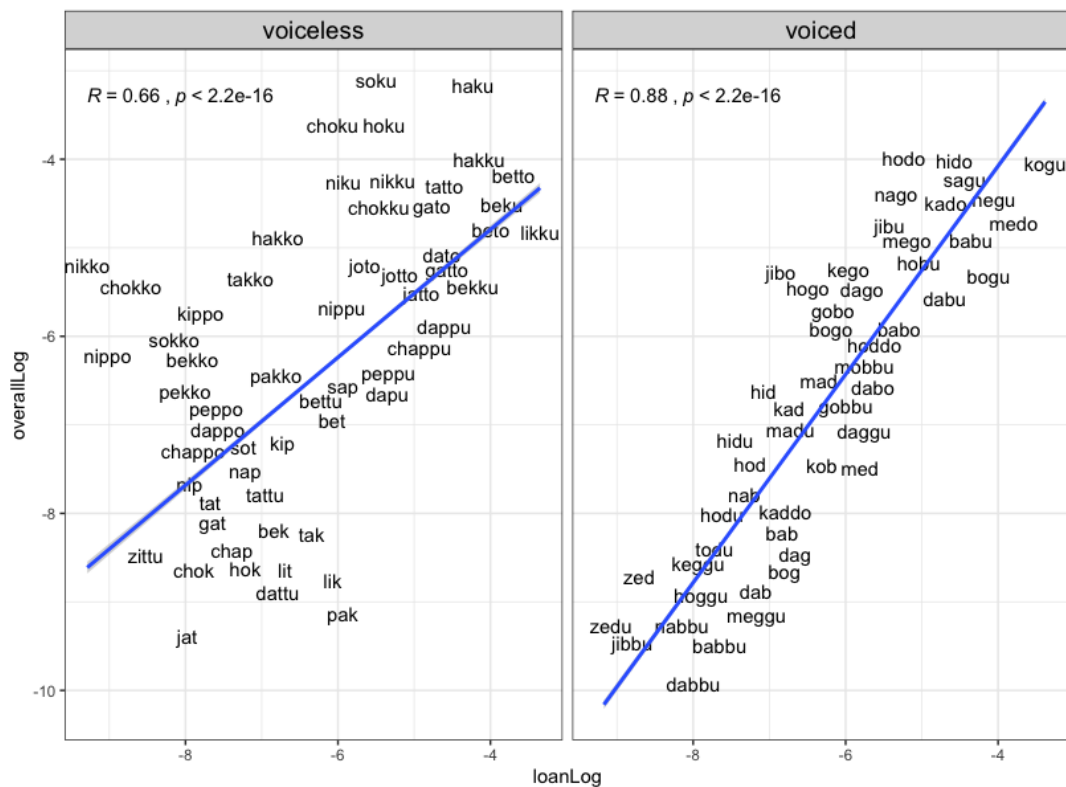


Figure 4.1 The correlation between the log-phonotactic scores for overall Japanese and that of loanwords in all stimulus words.

Taking into account the place of articulation, another Pearson’s test was run to determine the relationship between the phonotactic scores of overall Japanese and that of loanwords in three places of articulation for each voicing contexts. Figure 4.2 illustrates the correlation between the log-phonotactic scores for overall Japanese and that of loanwords in two voicing contexts. The top plots show the correlation in the voiceless context and the bottom plots show that of in the voiced context. The x-axes ‘loanLog’ stand for phonotactic scores of loanwords in Japanese, whereas the y-axes ‘overallLog’ stand for that of overall Japanese phonotactics. Symbols in the plots stand for each pronunciation type: ‘S’ = singleton, ‘G’ = gemination, ‘WE’ = ‘wrong epenthesis’, and ‘C’ = ‘CVC’. Note that the overlapping labels are avoided in the plots. Pearson’s correlation tests reveal that, for the voiceless context (top panels), the correlation was moderately positive in the labial ( $r = 0.59, p < .001$ ) and velar contexts ( $r = 0.53, p < .001$ ), whereas it was strong and positive in the alveolar context ( $r = 0.92, p < .001$ ). Overall, high loanword phonotactic scores show a tendency toward higher overall Japanese phonotactic scores. On the other hand, such tendency is not strong in voiceless labial and velar contexts. In the voiced contexts (bottom panels), the correlation was

statistically strong and positive between the two variables: labial, alveolar ( $r = 0.88, p < .001$ ), velar ( $r = 0.89, p < .001$ ). They are highly consistent regardless of place of articulation. In these contexts, increases in loanword phonotactic scores were correlated with increases in higher overall phonotactic scores.

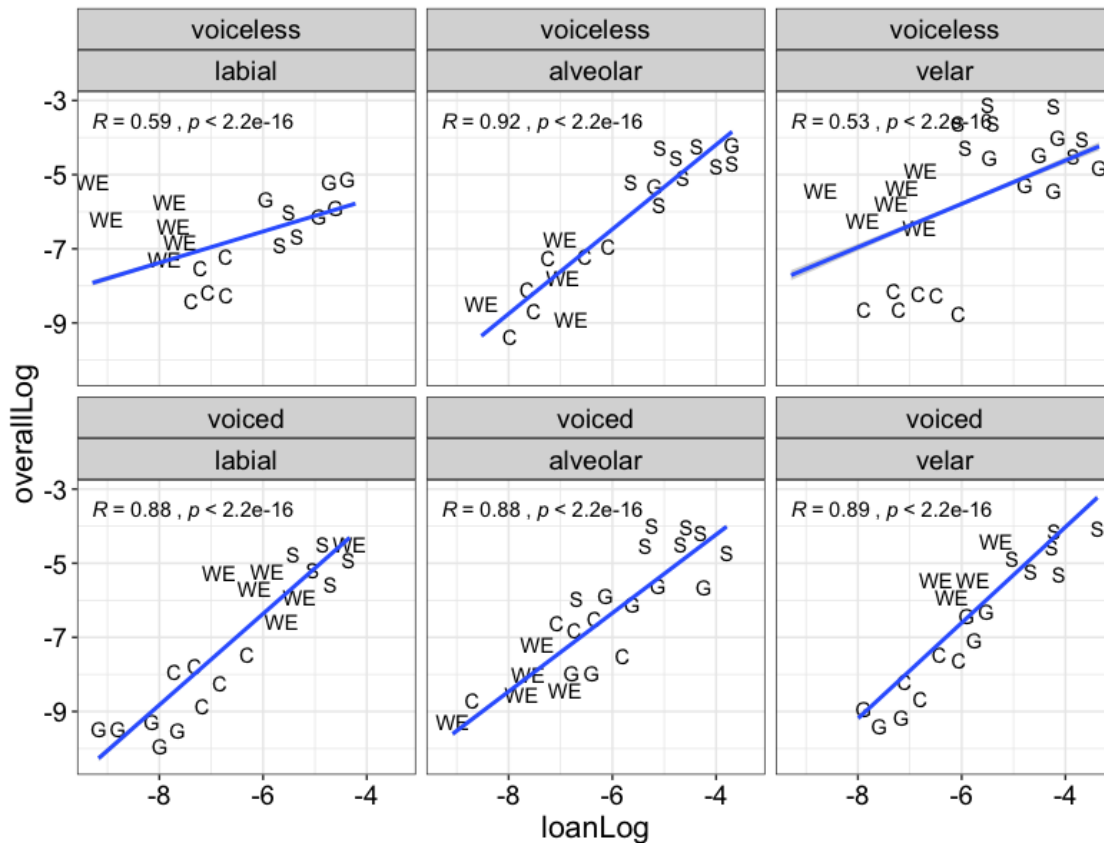


Figure 4.2 The correlation between the loanword log-transformed and overall log-transformed phonotactic probabilities for each place of articulations. Symbols in the plots stand for each pronunciation type: ‘S’ = singleton, ‘G’ = gemination, ‘WE’ = ‘wrong epenthesis’, and ‘C’ = ‘CVC’.

As can be seen in Figure 4.2, the strong correlation between loanword phonotactics and overall Japanese phonotactics in alveolar contexts seems to reflect the phonology of Japanese in which /tu/ and /du/ are illicit sequences. Overall, phonotactic scores of wrong epenthesis (e.g., /nago/) for voiced labial and velar contexts in loanwords are high as well as singletons (e.g., /hoku/) for voiceless labial and velar contexts in loanwords. Therefore, the statistical patterns in the lexicon might influence response patterns for loanword adaptation over loanword phonological rules. It would be shown by comparing response patterns in the following experiment to the phonotactic probability. As part of the analysis, we will investigate

whether loanword phonotactics and overall phonotactics make separate contributions to perceived well-formedness.

### 4.1.3 Procedure <sup>6</sup>

The same procedure as in the pilot study was used. To recap, the perceptual experiment was implemented as an online rating task. Participants were instructed to listen to the stimuli through headsets. First, they were asked to complete the audio system test. In the experimental phase, the stimulus word was presented orthographically (e.g., <pip>) to participants as they heard one of the pronunciations. After hearing the stimulus (e.g., *pippo*) participants were asked to judge whether the pronunciation they heard was how the word would be pronounced if this was a Japanese word, by rating how confident they are on a scale of 1-5 displayed on a computer screen. The scale is defined as follows: 1= “Confident that this is NOT how the word would be pronounced in Japanese”, 2= “Somewhat confident that this is NOT how the word would be pronounced in Japanese”, 3= “I really do not know if the word would be pronounced this way or not in Japanese”, 4= “Somewhat confident that this IS how the word would be pronounced in Japanese”, and 5= “Confident that this IS how the word would be pronounced in Japanese”. After the participant clicks one of the options and a ‘NEXT’ button, the next stimulus is presented. The audio stimulus would be played only once. After finishing the rating task, the participants filled out a questionnaire. There was no training phase or practice phase. No feedback was given during the experiment. The entire procedure was described in the relevant native language of the participants. The whole experiment lasted approximately 30 minutes.

### 4.1.4 Questionnaires

To investigate the extent to which each speaker has been exposed to Japanese throughout their life, other than taking a Japanese course at university, and how familiar with Japan they are, all participants were asked to complete a questionnaire at the end of the experiment. The questionnaire was designed to determine associations between experimental results and extra linguistic factors (Appendix D). Specifically, the participants’ relatedness and exposure to the Japanese language and aspects of the Japanese culture were focused on. This is based on an assumption that participants who are more involved in Japanese social practices might assess the target stimuli accurately even without possessing an established Japanese lexicon.

---

<sup>6</sup> This experiment was preregistered using AsPredicted (<https://aspredicted.org/index.php>). The ID is #15007. The preregistration document appears in Appendix C.

Questions are adapted from other studies (First People's Cultural Council, 2013), and they were modified for the current study. The questionnaire was composed of three sections: general information, degree of knowledge about Japanese, and degree of exposure to Japanese. Thus, in addition to the extra-linguistic factors, the individuals' language proficiency was measured through the questionnaire as well. Participants were asked to self-evaluate their language ability. This is in response to comments from a Japanese lecturer of the Japanese programme (personal communication, April 4, 2018), that Japanese proficiency varies between students especially in a first-year course as some students study Japanese at high school, whereas other students are absolute beginners. Individual students notice their own proficiency level in class, especially regarding speaking. Therefore, the questionnaire contains three criteria of proficiency: speaking, comprehension, and reading. The questionnaire was filled out after data collection, using an online survey tool.

Responses in the questionnaire were measured using a Likert (Likert, 1932) scale which the ends of continuum are presented with opposite values (e.g., always/never). Participants were required to indicate their degree of agreement with each statement by clicking at one of five or six options. For degree of knowledge about Japanese, one end was "very well" and the other end was "not at all". For degree of exposure to Japanese, one end of the continuum was "daily" or "always" (depending on questions) and the other end is "never".

#### **4.1.5 Principal component analysis (PCA)**

After collecting the responses of the questionnaires, all responses were converted to numeric values except the binary questions. Then, they were statistically analysed using the technique of principle component analysis (hereafter, PCA) which has previously been used for analysing socio-linguistic data collected by questionnaires (e.g., Flege, 2009; Hashimoto, 2019; Wang, 2017). PCA is a statistical method of dimensionality reduction by which a number of original explanatory variables are transformed into a smaller set of linearly uncorrelated variables called components/factors (Dunteman, 1989). This is because many observed variables might be correlated with one another. The purpose of using PCA in this study is to reduce variables in questionnaire in order to interpret and capture extra-linguistic factors that influence the responses of participants.

Questions for exposure to Japanese in the questionnaire completed by learners and non-learners of Japanese are the same. However, questions for knowledge of Japanese are slightly different between these two groups. In addition, questions for native Japanese speakers are totally different than those of the other two groups. The purpose of a questionnaire for native

Japanese speakers is to obtain general demographic information of participants. Therefore, PCA was carried out separately for each group.

There are many statistical methods in terms of how many factors should be extracted and how they will be extracted (Beavers, Lounsbury, Richards, & Huck, 2013). I consider following three criteria:

1. Kaiser's criterion of 1: eigenvalues  $> 1$
2. Cumulative percentage of variance
3. Scree plot test with parallel analysis: A graphical method in which number of actors would be indicated by a break in the graph.

PCA was carried out using the principal () function in the *psych* package in R (Revelle, 2015).

#### **4.1.5.1 PCA of native Japanese speakers**

PCA was conducted on the six items with oblique rotation (oblimin) using the *psych* package in R. The Kaiser-Meyer-Olkin (hereafter KMO) measure verified the sampling adequacy for the analysis  $KMO = .65$  (accepting values is over .5 according to Kaiser, 1974), and all KMO values for individual items were  $>.58$ , which is above the acceptable limit of .5, except (a question: grade year) = 0.27. Therefore, this question was excluded in this phase. For five items and 20 participants, Bartlett's test of sphericity,  $X^2(10) = 51.81037$ ,  $p < .001$ , indicated that correlation between items were sufficiently large for PCA. In the final form of analysis KMO was .69 and all individual items were  $>.63$ .

An initial analysis was run to obtain eigenvalues for each component in the data. Two components had eigenvalues over Kaiser's criterion of 1 and in combination explained 83% of the variance. Given Kaiser's criterion, two components should be extracted. However, there are many statistical methods to determine how many components should be extracted. Therefore, another statistical method, the scree plot with parallel analysis was conducted. The scree plot in Figure 4.3 shows the eigenvalues on the y-axis and the number of components on the x-axis. The cut-off point for selecting components should be at the inflection point where the slope of the curve (i.e., blue line) is a sharp drop (Field, Miles, & Field, 2012). The same description applies hereafter for all of the scree plots below. This cut-off point is indicated by the red dotted line and showed inflexions that would justify retaining one component.

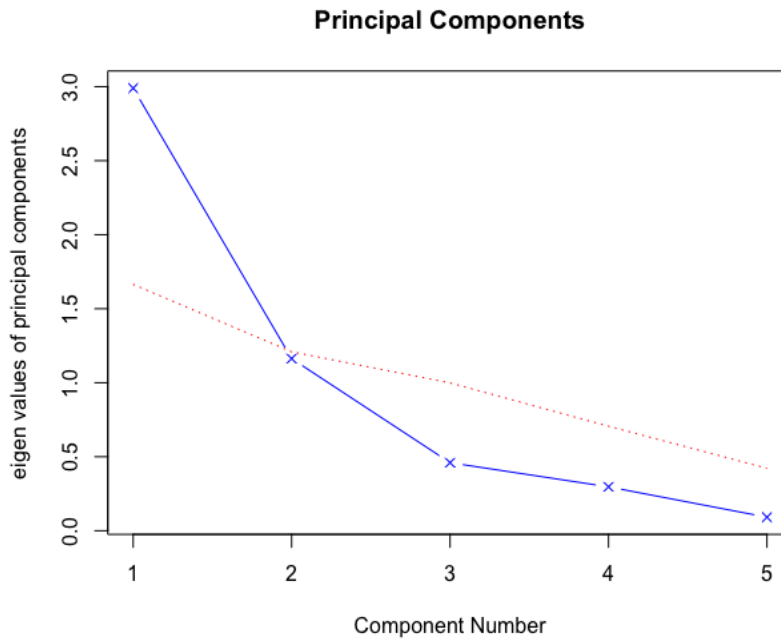


Figure 4.3 Scree plot from PCA of native Japanese speakers. While the blue line indicates the number of generated components, the red dotted line indicates the cut-off point.

After taking the results, sampling size and questions in components into consideration, two components were retained in the final analysis. This is because questions in PC1 and PC2 show different factors. Table 4.3 show the factor loadings after rotation.

Table 4.3 Two principal components revealed by PCA for native Japanese speakers

PC-English (PC1)	How well can you understand/read English?	0.95
	Approximately how many years have you been in New Zealand?	0.93
	How well can you speak English?	0.80
	Languages in which you can have an everyday conversation	0.33
PC-Language (PC2)	If you are studying Linguistics, at what level?	0.93
	Languages in which you can have an everyday conversation	0.72

The items that cluster with the same components suggest that the first component (PC1) seems to be related to exposure to English in NZ. More specifically, the questions in PC1 were about self-reported levels of proficiency in English and period of stay in NZ. The second component (PC2) seems to be related to exposure to other languages in NZ and overall knowledge of languages. Thus, considering questions in the two components, PC1 was labelled as “PC-English” and the second component was labelled as “PC-Language”. Then, these extra-

linguistic scores were assigned to each participant. A higher number indicates higher exposure. These scores were used to assess the tendency of responses for participants accordingly.

#### **4.1.5.2 PCA of NZE-speaking learners of Japanese**

All questions about the degree of knowledge of Japanese, and degree of exposure to Japanese were included, in addition to questions about the length of studying Japanese and experience of visiting Japan. A question about reading proficiency was used to obtain a general idea about the extent to which participants can read in Japanese and/or English. Therefore, the question was not included in the PCA from the outset. First, Bartlett's test was run for 22 questions to judge whether questions had correlations with each other. The  $p$  value in the results has to be a significant  $<0.05$ . In the initial run,  $p$  values were not obtained. This is because the sample size was small for a number of variables (i.e., questions). Therefore, one question had to be removed from the data analysis. Which question to select used variable importance values with Random Forest<sup>7</sup> (Breiman, 2001) using the *ranger* package (Wright & Ziegler, 2015) and the *randomForest* package (Liaw & Wiener, 2002). Then, a question whose variable importance value was lowest is excluded. This procedure was for avoiding subjective judgements. Then, Bartlett's test was run for 21 questions. Then, a KMO test (Kaiser-Meyer Olkin test) was run to observe whether the sample size is appropriate. The KMO statistic varies between 0 and 1. Kaiser (1974) recommends accepting values greater than .5 as barely acceptable. Variables whose value was lower than .5 was removed from the analysis and this process was done one-by-one.

Finally, PCA was conducted with oblique rotation (oblimin) and Bartlett's test of sphericity,  $X^2(36) = 110.75$  ( $p < .001$ ), for nine items and 22 participants, which indicated that correlation between items were adequate for PCA. The analysis KMO = .77 and all individual items were  $>.73$ . An initial analysis was run to obtain eigenvalues for each component in the data. Two components had eigenvalues over Kaiser's criterion of 1 and in combination explained 68% of the variance. Given Kaiser's criterion, two components should be extracted. In relation to the first group, native speakers of Japanese, the scree plot showed inflexions that would justify retaining two components (Figure 4.4). Therefore, two components were extracted for this group.

---

<sup>7</sup> Random forest is a statistical machine-learning approach to examine the importance of multiple predictors.

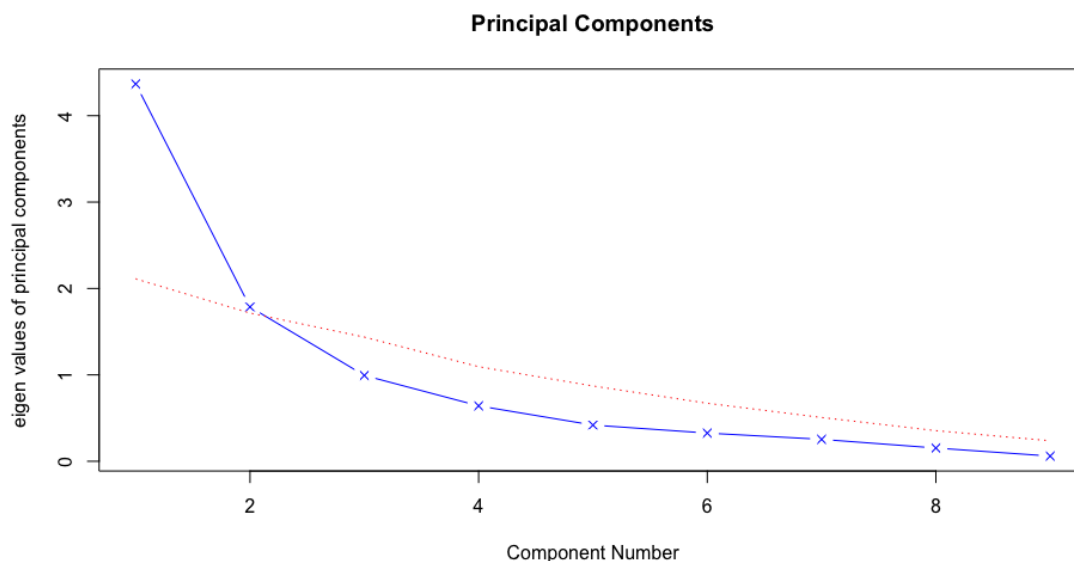


Figure 4.4 Scree plot from PCA of L2 learners of Japanese. While the blue line indicates the number of generated components, the red dotted line indicates the cut-off point.

The two principal components (factors) revealed by PCA are shown in Table 4.4. From the questions, it can be seen that the first component (PC1) was related to exposure to Japanese in Japan, and the second component (PC2) was related to exposure to Japanese in NZ. Thus, PC1 was labelled as “Exposure to Japanese in Japan: PC-JJ” and the second as “Exposure to Japanese in NZ: PC-JNZ”.

Table 4.4 Two principal components revealed by PCA for learners of Japanese

(PC1) PC-JJ “exposure to Japanese in Japan”	How long have you been in Japan?	0.89
	What was the purpose of the visit?	0.89
	How many times have you been to Japan?	0.87
	When did you last visit Japan?	0.72
(PC2) PC-JNZ “exposure to Japanese in NZ”	How often do you access websites about modern Japanese culture?	0.89
	How often do you read/browse Japanese written cartoons or magazines?	0.88
	How often do you watch Japanese TV programs or Japanese movies on the internet (with English subtitles)?	0.61

It is possible that some questions are not correlated with others, so do not rate highly on the PCA, and yet still predict response patterns well. To double check this possibility, I conducted a random forest (Breiman, 2001) including the PCs, the main predictors and all other



questions, predicting responses by using the *randomForest* package (Liaw & Wiener, 2002) the *ranger* package (Wright & Ziegler, 2015) in R, as we are particularly interested in this group in the current study. The result revealed that PC1 and PC2 are more important variables than questions that ask about individuals' knowledge of Japanese (e.g., Understanding Japanese, Speaking Japanese or Course Level) and length studying Japanese. These questions are not derived as a PCA factor.

#### **4.1.5.3 PCA of NZE-speaking non-learners of Japanese**

There are 20 participants in all. Three participants reported that they studied Japanese at high school but for a duration of less than six months each. All participants reported that they understand no more than a few words in Japanese. For this group, none of the participants have visited Japan. Therefore, questions related to visiting Japan were not included in the data analysis. All questions about degree of knowledge of Japanese, and degree of exposure to Japanese were included. The same process of extracting components was conducted on the eight items with oblique rotation (oblimin). The procedure to select the eight items was the same as the previous group. The KMO measure verified the sampling adequacy for the analysis  $KMO = .73$ , and KMO values for individual eight questions were  $>.64$ , which is above the acceptable limit of  $.5$ . For eight items and 20 participants, Bartlett's test of sphericity,  $X^2(28) = 16373$ ,  $p < .001$ , indicated that correlation between items were adequate for PCA. An initial analysis was run to obtain eigenvalues for each component in the data. One component had eigenvalues over Kaiser's criterion of 1 and the component explained 69% of the variance. The scree plot showed inflexions that would justify retaining one component, as well (Figure 4.5). Therefore, one component was extracted for this group.

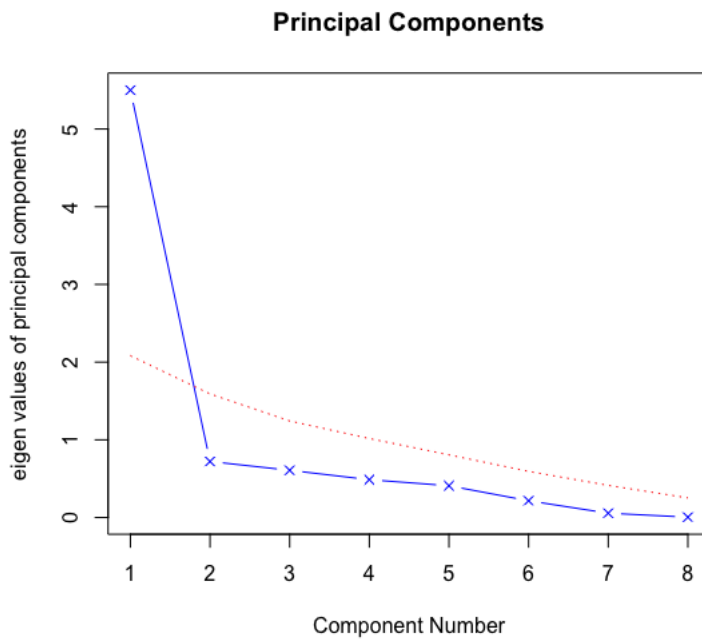


Figure 4.5 Scree plot from PCA of non-learners of Japanese. While the blue line indicates the number of generated components, the red dotted line indicates the cut-off point.

The one principal component (factor) revealed by PCA are shown in Table 4.5. The component was related to exposure to Japanese in NZ. Thus, PC1 was labelled as ‘Exposure to Japanese in NZ: PC-JNZ.

Table 4.5 One principal component revealed by PCA for non-learners of Japanese

(PC1) PC-JNZ “exposure to Japanese in NZ”	How often do you watch Japanese TV programs or Japanese movies on the internet (with English subtitles)?	0.95
	How often do you access websites about modern Japanese culture?	0.94
	How often do you read/browse Japanese written/drawn cartoons or magazines?	0.93
	How often do you access websites that contain Japanese language resources?	0.84
	How often do you access websites about traditional Japanese culture?	0.84
	How often do you hear Japanese outside of University at the present time?	0.73
	I know how to say some basic phrase (e.g., My names is....)	0.73
	How well can you understand Japanese?	0.61

## 4.2 Phonotactic Effects

### 4.2.1 Statistical analysis

In this section, the analysis of predicted response focuses on the relationship with statistical phonotactic scores, as based on the dictionary dataset extracted from the CSJ as described earlier in this chapter (§4.1.2.3). As part of the analysis, this thesis investigates whether loanword phonotactics and overall phonotactics make separate contributions to perceived well-formedness. As a reminder, there are two categories of log-phonotactic scores: loanword phonotactic scores and overall Japanese phonotactic scores. Regarding **RQ4**, when participants do not have knowledge of loanword phonology, they might access available statistical phonotactics. Therefore, an analysis particularly probing this question was explored. Since loanwords are a subset of the Japanese lexicon, each phonotactic score is expected to share similarities. In order to deal with collinearity, a predictor residualisation approach was used in which predictor variables (i.e., phonotactic scores) are residualised (e.g., Hendrix, Bolger, & Baayen, 2017; Kuperman, Bertram, & Baayen, 2010; Soskuthy & Hay, 2017). The dependent variable was response 1-5 Likert scales in the experiment. The independent variables are logged phonotactic scores for loanwords and overall Japanese.

First, using R (R Core Team, 2018), I ran a regression analysis using  $X_2$  = ‘loanword scores’ to predict  $X_1$  ‘overall Japanese scores’. Then, a new variable  $X_3$  was created from the residuals of the regression analysis (i.e.,  $\text{response} \sim X_1 + X_3$ ). Response patterns in the experiment (i.e., participants’ ratings) were predicted to be different between voiced and voiceless stimuli across groups, thus mixed effect models were run for each category for native Japanese speakers, learners of Japanese, and non-learners of Japanese. The models have responses (i.e., ratings) as the dependent variable and the fixed effects of loanword phonotactic scores and residualised overall Japanese phonotactic scores with SUBJECT as a random intercept. For loanword phonotactics, if participants rely on loanword phonotactics, words with a high score show a tendency toward higher rating. This will be illustrated by positive slopes, whereas negative slopes indicate less likelihood of using loanword phonotactics. The same applies to overall Japanese phonotactics as well.

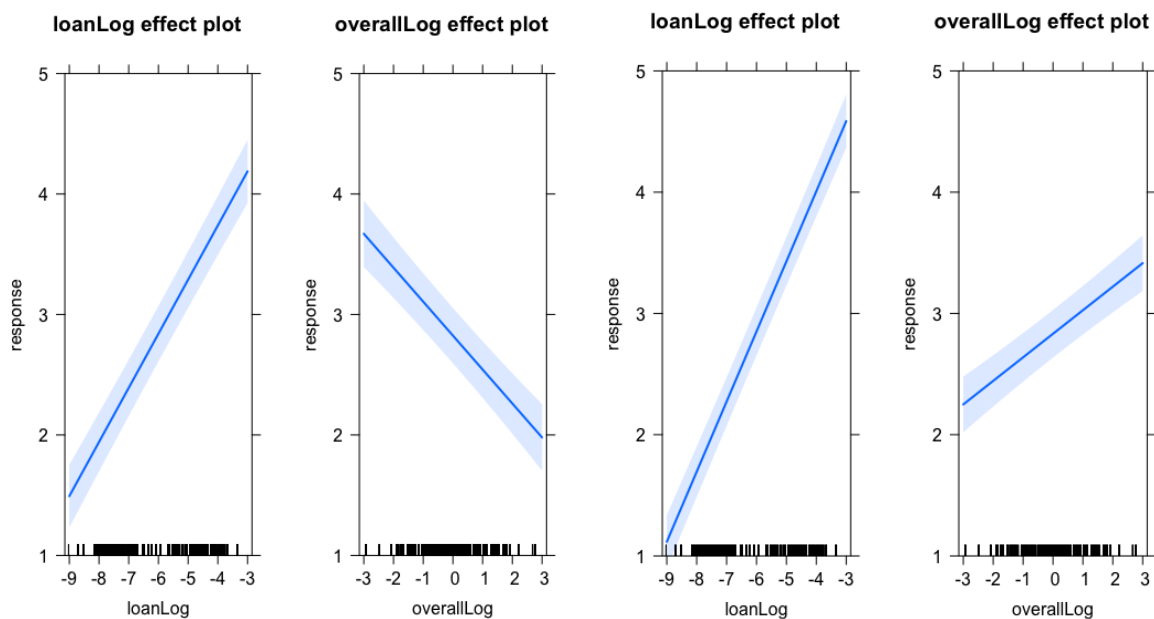
### 4.2.2 Results

Figure 4.6 illustrates the results of the voiceless context for native Japanese speaker in (a), L2 learners of Japanese in (b) and non-learners of Japanese in (c). For native Japanese speakers (Figure 4.6a), the effect of loanword phonotactic scores show a significant positive result ( $p <$

0.001), which suggests that words with higher loanword phonotactic scores are more likely to receive higher ratings. On the other hand, overall Japanese phonotactic scores have a significantly negative impact on ratings ( $p < 0.001$ ). As overall Japanese phonotactic scores increase, ratings become more negative. This could be attributed to the fact that the Japanese speakers did not confidently reject CVC tokens and wrong epenthesis with [tu] and [du] as illicit pronunciations in Japanese but these CV sounds occur in loanwords (details will be shown in the next section). More particularly, the overall phonotactic score captures only the overall phonotactic variation in the stimuli that is not linked to loanword phonotactics. This would be a sign that it's not a good loanword, so it is coherent for fluent raters to treat this negatively.

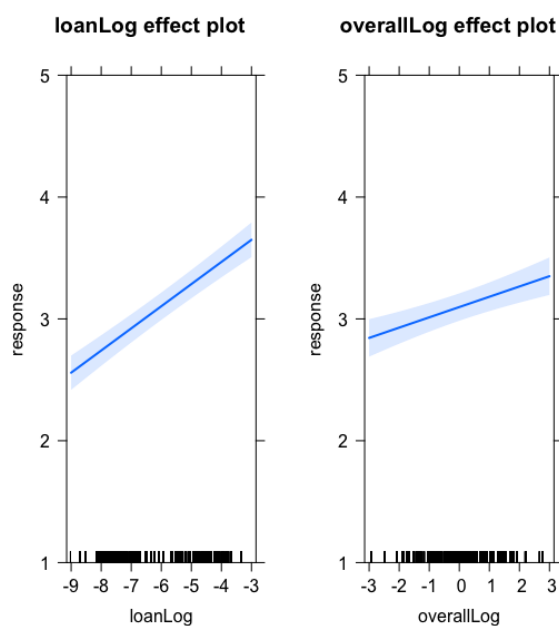
For L2 learners of Japanese (Figure 4.6b), both the effects of loanword phonotactic scores and overall Japanese phonotactic scores show significant and positive results ( $p < 0.001$ ). As both phonotactic scores increase, the ratings become higher. However, the slopes of the figures suggest that learners of Japanese are more likely to access to the loanword phonotactics than overall Japanese phonotactics.

For non-learners of Japanese (Figure 4.6c), both the effects of loanword phonotactic scores and overall Japanese phonotactic scores show significant and positive results ( $p < 0.001$ ). As both phonotactic scores increase, the ratings become higher. The slopes of the figures suggest that although non-learners are more likely to access to the loanword phonotactics than overall Japanese phonotactics, the degree of accessing was quite different from other two groups.



(a) Japanese: Voiceless Stimuli

(b) L2 learners: Voiceless Stimuli

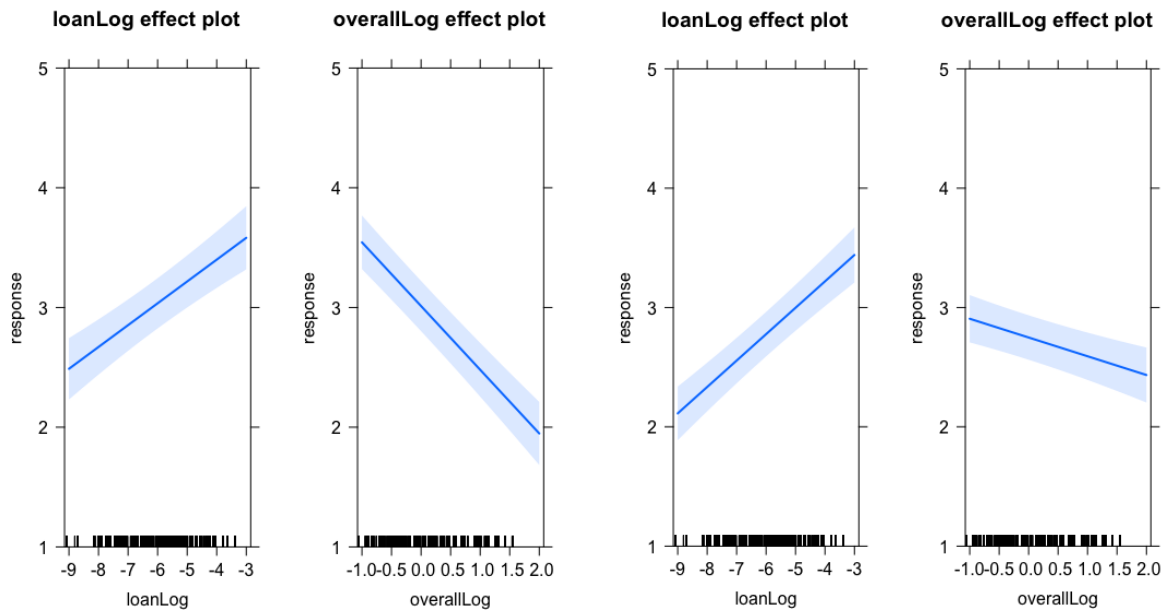


(c) Non-learners: Voiceless Stimuli

Figure 4.6 The effects of phonotactic scores on responses for each group: voiceless stimuli.

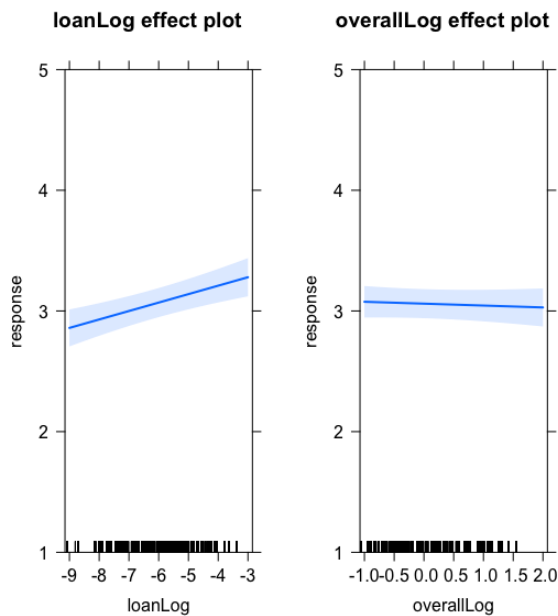
Next, the results of the voiced stimuli were considered. Figure 4.7 illustrates the results of the voiced context for native Japanese speakers (Figure 4.7a), learners of Japanese (Figure 4.7b), and non-learners (Figure 4.7c). In this context, for both native Japanese speakers and L2 learners, the effect of loanword phonotactic scores shows a significant positive result ( $p < 0.001$ ), which suggests that words with higher loanword phonotactic scores are more likely to receive a higher rating. On the other hand, overall Japanese phonotactic scores have a significantly negative impact on ratings in native Japanese speakers ( $p < 0.001$ ) and L2 learners ( $p < 0.05$ ). This is due to voiced geminates: [bb], [dd], [gg] are licit phonotactic sequences for loanwords in Japanese while they are illicit in native Japanese. The effects of loanword phonotactics is weaker in both groups in comparison to the voiceless context, but this could be due to different reasons. While native Japanese speakers rated CVC higher, L2 learners rated gemination higher even though geminates are less common in the labial and velar contexts (details will be shown in the next section). Both groups were rating words with high overall phonotactic scores low. An explanation for this finding could be due to the presence of the stimuli ‘wrong epenthesis’ in voiced labial and velar that have high overall phonotactic scores (see §4.1.2.3; Figure 4.2). The listeners judged that these stimulus words would not be pronounced with wrong epenthesis in loanwords in Japanese.

For non-learners (Figure 4.7c), the effects of loanword phonotactic scores showed significant and positive results ( $p < 0.001$ ) but overall Japanese phonotactic scores did not show significance ( $p = 0.54$ ). Most importantly, likelihood of ratings between learners and non-learners are quite different, indicating learners acquire Japanese sublexicon phonology to some extent.



(a) Japanese: Voiced Stimuli

(b) L2 Learners: Voiced Stimuli



(c) Non-learners: Voiced Stimuli

Figure 4.7 The effects of phonotactic scores on responses for each group: voiced stimuli.

Overall, the results suggest that both native speakers of Japanese and learners of Japanese had access to loanword phonotactics during the perceptual experiment. In addition, native speakers of Japanese tend to be sensitive to overall Japanese phonotactics. Thus, results from this experiment provide some evidence for that loanword phonotactics and overall phonotactics make separate contributions to perceived well-formedness.

### 4.3 Statistical Analysis for Phonological Processing

#### 4.3.1 Factors

In order to examine the participants’ knowledge of sublexicon phonology, some factors were considered in relation to research questions. The dependent variable was response 1-5 Likert scales in the experiment. The response categories in Likert scales have a rank order which is original scales. That is, the data are not continuous and have equal intervals between values in categories. Some might ask whether using a parametric test is inappropriate for ordinal data, however, Norman (2010) and Kizach (2014) recommend using linear regression Mixed-Effects Models for Likert ratings. Following previous linguistic studies using mixed-effect models for analysing the data from Likert ratings (Gibson, Piantadosi, & Fedorenko, 2011; Schmidt, Janse, & Scharenborg, 2016), I use Likert scales as a dependent variable in regression with an assumption that intervals between values are equal.

There were two broad categories for the independent factors: phonological and extra-linguistic factors. Independent factors on responses are summarised in Table 4.6. Independent variables are VOICING TYPE (2 levels: voiced/voiceless), PRONUNCIATION TYPE (4 levels: singleton/gemination/wrong epenthesis/CVC), POA (3 levels: labial/alveolar/velar), GROUP (3 levels: Japanese/learners/non-learners), factors by PCA, and trial number. Trial number was standardised by converting the variable to a z-score.

Table 4.6 Factors considered in analysis of dataset

	<b>Factor</b>	<b>Levels</b>
Phonological Factors	VOICING TYPE	voiced/voiceless
	PRONUNCIATION TYPE	singleton/gemination/wrong epenthesis/CVC
	POA	labial/alveolar/velar
Extra-Linguistic Factors	GROUP	Japanese/learners/non-learners
	PC-JJ (Exposure to Japanese in Japan)	Index score
	PC-JNZ (Exposure to Japanese in NZ)	Index score
	PC-English (Exposure to English in NZ)	Index score (applicable to only Japanese)
other	PC-Language (Exposure to other languages in NZ)	Index score (applicable to only Japanese)
	zTrial (z-scored 1-240 trial numbers)	Index score

#### **4.3.1.1 Phonological factors**

##### ***Voicing type***

In nativised loanwords, word-final voiceless stop consonants (after a lax vowel in a donor language) are consistently geminated as shown in the data from the BCCWJ and in other studies (Katayama, 1998; Shirai, 1999). Voiced stops are not geminated as frequently as voiceless stops in the same contexts based on stochastic data. In addition, occurrence of gemination varies between place of articulation in voiced contexts. This study predicts that L2 learners with high exposure to Japanese would be expected to be able to acquire fine-grained knowledge regarding the effects of voicing. Thus, when word-final stops are voiceless, listeners with knowledge of Japanese prefer gemination. On the other hand, singleton would be preferred when word-final stops are voiced.

##### ***Pronunciation type***

There are four pronunciation types for the confidence-rating task: singleton, gemination, wrong epenthesis and CVC. The predictions are that participants with more exposure to Japanese would give higher ratings to gemination and singleton, which do not violate Japanese phonological rules. In addition, wrong epenthesis should receive lower ratings, if participants have a good knowledge of the Japanese phonological convention that final consonants in word-final position are disallowed and must be repaired through appropriate epenthetic vowels.

##### ***POA***

Place of articulation in coda position is included as a factor, which has three levels: labial, alveolar, and velar. POA is expected to influence the realisation of word-final stops in voiced contexts. The alveolar stop [d] tends to geminate more frequently than labial and velar stops. Thus, gemination in the voiced context can be expected in the following order: [dd] > [gg] > [bb]. This hierarchical model is only relevant when gemination is selected over other pronunciation types. Thus, participants with more knowledge of Japanese will behave differently according to POA.

#### **4.3.1.2 Extra-linguistic factors**

##### ***Group***

There are three levels: Japanese (i.e., native speakers of Japanese), learners (i.e., NZE-speaking L2 learners of Japanese) and non-learners (i.e., NZE-speaking monolinguals).



### *PCA scores*

I assume that rating values for pronunciation types reflect each participant's exposure to the Japanese language.

#### **4.3.2 Mixed effects models for phonological process (i.e., pronunciation type)**

All responses in the experiment were analysed using the *lme4* package in R (Bates et al., 2015; R Core Team, 2018) to perform a linear generalised mixed effects analysis of the relationship between pronunciation type and voicing type and groups. In order to measure performance across participants, firstly mean reaction time in the perception task was analysed. There were 14,876 responses in total. Then, the distribution of the data was checked and reaction times more than two standard deviations above the mean were removed as outliers from further analysis. It is possible that responses with long reaction times are real responses and not outliers, however, long responses might also reflect loss of attention or distraction. Therefore, in the current study, 0.33% of observations were removed, and 14,827 observations remained for analysing the responses shown in Table 4.7.

Table 4.7 Numbers of stimuli of each place of articulation by group

Group	# of participants	Items			Total of Items
		Labial	Alveolar	Velar	
Japanese	20	1593	1595	1594	4782
Learners	22	1749	1750	1757	5256
Non-learners	20	1596	1599	1594	4789
Total	62	4938	4944	4945	14,827

The statistical models that predict the responses in the current study are created in the following manner. First, a full model with fully crossed and specified random effects structure was created. Then, if the inclusion of a slope led to convergence errors, the slope that contributes least to the model is dropped in order to obtain convergence. Using stepwise regression, models are compared to each other using an ANOVA () function to see which fits the best, based on the Akaike Information Criterion (AIC) and considering  $p$  values (Baayen, Davidson, & Bates, 2008; Jaeger, 2008). If the ANOVA test shows a significant ( $p < .05$ ) improvement in model fit, the factor is retained. Non-significant main effect is also removed from the model as well (Barr, Levy, Scheepers, & Tily, 2013).

In order to avoid multicollinearity, the diagnostic Variance Inflation Factor (VIF) was used to test whether any two factors were highly correlated. A VIF score was calculated for each factor used in the model, and all VIF scores were less than 9. It seems a little high, however, the maximum acceptable level of VIF is less than 10 (Hair, Black, Babin, Anderson, & Tatham, 1998). Predictions for the well-formedness judgement are listed in Table 4.8.

Table 4.8 Predictions for well-formedness

Phonological Context	Variable	Expected rating
		<b>high &gt; low for well-formedness</b>
Voiced	Effect of pronunciation	singleton > gemination > wrong epenthesis > CVC
Voiceless		gemination > singleton > wrong epenthesis > CVC
Voiced for gemination	Effect of POA	alveolar > velar > labial
Well-formedness	Language Group	Japanese will give rating more confident with loanword phonology than learners

## 4.4 Results

### 4.4.1 Phonological process: Full dataset

All 14,827 responses from 62 participants across three groups were fit into a generalised linear mixed effects model. The below is the best model for the full dataset. The model contains SUBJECT and WORD as random intercepts.

`lmer (response ~ VOICING_TYPE * PRONUNCIATION_TYPE * GROUP +  
(1 |SUBJECT) + (1 |WORD), data = data, REML = F)`

The results of the model are presented in Table 4.9. The reference level is voiced singleton of Japanese participants and the predicted rating is 3.84. There is a significant three-way interaction between VOICING TYPE, PRONUNCIATION TYPE and GROUP. This is a result of different types of pronunciation being preferred for different voicing types and for the groups.

Table 4.9 Model summary for full dataset (all groups)

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	3.84	0.10440	36.793	< 2e-16
VOICING_TYPE voiceless	-1.10700	0.07287	-15.188	< 2e-16
PRONUNCIATION_TYPE gemination	-0.65960	0.06985	-9.443	< 2e-16
PRONUNCIATION_TYPE wrong epenthesis	-2.05800	0.06982	-29.47	< 2e-16

PRONUNCIATION_TYPECVC	-0.60150	0.06976	-8.622	< 2e-16
GROUP learners	-0.25500	0.14270	-1.787	0.07707
GROUP non-learners	-0.48000	0.14610	-3.286	0.00142
VOICING_TYPE voiceless: PRONUNCIATION_TYPE gemination	2.08700	0.09870	21.148	< 2e-16
VOICING_TYPE voiceless: PRONUNCIATION_TYPE wrong epenthesis	0.87720	0.09876	8.883	< 2e-16
VOICING_TYPE voiceless: PRONUNCIATION_TYPE CVC	0.71020	0.09864	7.2	6.32E-13
VOICING_TYPE voiceless: GROUP learners	1.22700	0.09639	12.728	< 2e-16
VOICING_TYPE voiceless: GROUP non-learners	1.15600	0.09861	11.72	< 2e-16
PRONUNCIATION_TYPE gemination: GROUP learners	0.69660	0.09650	7.218	5.52E-13
PRONUNCIATION_TYPE wrong epenthesis: GROUP learners	0.61710	0.09648	6.396	1.64E-10
PRONUNCIATION_TYPE CVC: GROUP learners	-1.34500	0.09647	-13.942	< 2e-16
PRONUNCIATION_TYPE gemination: GROUP non-learners	0.61790	0.09868	6.262	3.90E-10
PRONUNCIATION_TYPE wrong epenthesis: GROUP non-learners	1.69100	0.09874	17.125	< 2e-16
PRONUNCIATION_TYPE CVC: GROUP non-learners	-0.19100	0.09868	-1.936	0.0529
VOICING_TYPE voiceless: PRONUNCIATION_TYPE gemination: GROUP learners	-1.98100	0.13640	-14.525	< 2e-16
VOICING_TYPE voiceless: PRONUNCIATION_TYPE wrong epenthesis: GROUP learners	-1.13900	0.13650	-8.348	< 2e-16
VOICING_TYPE voiceless: PRONUNCIATION_TYPE CVC: GROUP learners	-0.70530	0.13640	-5.17	2.37E-07
VOICING_TYPE voiceless: PRONUNCIATION_TYPE gemination: GROUP non-learners	-2.06000	0.13950	-14.762	< 2e-16
VOICING_TYPE voiceless: PRONUNCIATION_TYPE wrong epenthesis: GROUP non-learners	-0.96740	0.13960	-6.928	4.45E-12
VOICING_TYPE voiceless: PRONUNCIATION_TYPE CVC: GROUP non-learners	-0.69410	0.13950	-4.975	6.61E-07

Plots of the three-way interaction between VOICING TYPE, PRONUNCIATION TYPE and GROUP are shown in Figure 4.8. The left plot shows the interaction between pronunciation type and voicing type for native speakers of Japanese, the middle plot shows the same interaction for L2 learners, and the right slot represents non-learners. The x-axes are the two voicing types and the y-axis is the predicted response rating. The four lines stand for four pronunciations, respectively. The light-blue solid lines represent singleton, the pink dashed lines represent gemination, the green dotted lines represent wrong epenthesis, and the red dash-dotted lines represent CVC.

## VOICING\_TYPE\*PRONUNCIATION\_TYPE\*GROUP effect plot

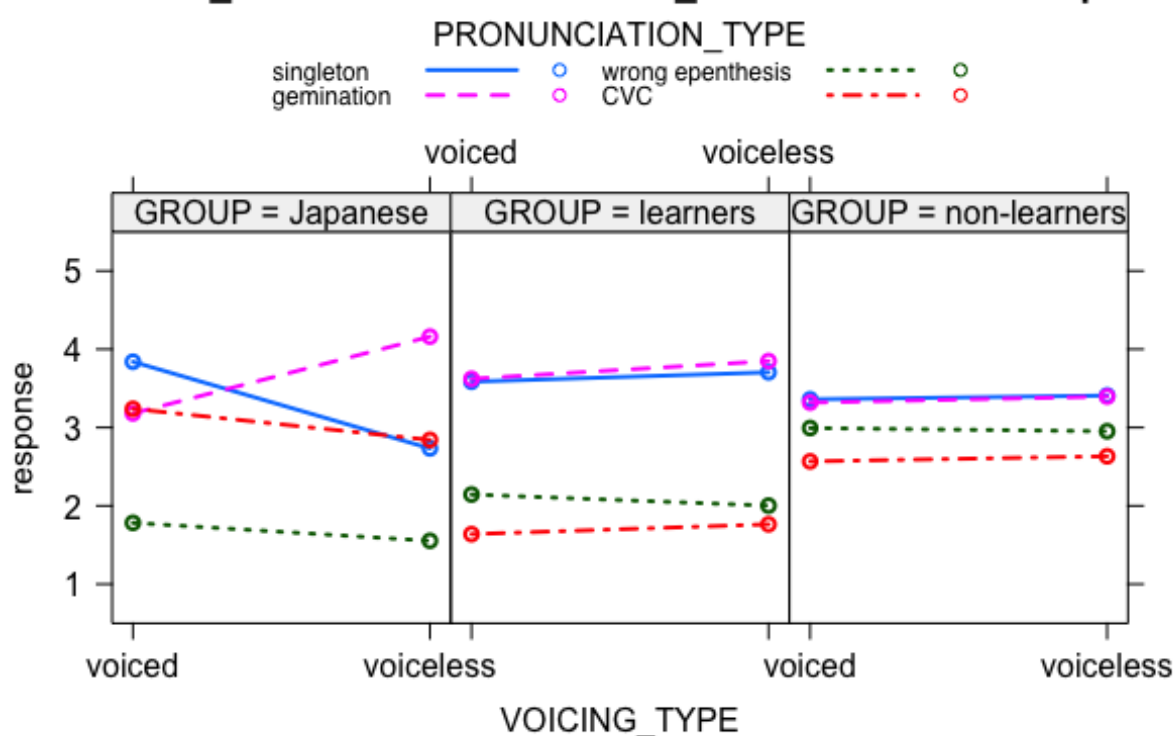


Figure 4.8 Plots of the three-way interaction between VOICING TYPE, PRONUNCIATION TYPE and GROUP in the full dataset.

The interaction effect of voicing type is observed in Figure 4.8. For native Japanese speakers, the difference between voiced and voiceless stimuli is larger for gemination and singleton. Singleton is more likely to be preferred in the voiced context and gemination is rated higher than singleton in a voiceless context. Put another way, the response patterns of gemination and singleton are relatively predictable by voicing types. However, this is not the case for non-native speakers of Japanese. For learners, the difference between singleton and gemination seems to be slightly detected in the voiceless context but not in the voiced context. For non-learners there is no difference between singleton and gemination for both voicing types. This suggests that native speakers of Japanese are more sensitive to voicing types when judging well-formedness between gemination and singleton but L2 learners are not. Overall, CVC and wrong epenthesis tend to be disfavoured across groups, and while this pattern is clearly observed for learners of Japanese, native Japanese speakers rated higher for CVC than learners. This result will be discussed when datasets are analysed separately for each group.

Interestingly, overall response patterns in learners and non-learners are very similar. For both groups, CVC and wrong epenthesis tend to receive lower ratings than gemination and singleton regardless of the voicing types. These results indicate that the pattern shown for

pronunciation type is comparable for both groups. In addition, the effect of voicing type underlying these patterns is very similar regardless of whether respondents have studied Japanese or not. Since the three-way interaction was detected, in order to explore these patterns further, subset data analyses were performed.

#### 4.4.2 Phonological process: Differences between native Japanese speakers and learners of Japanese

In order to examine the effects of PRONUNCIATION TYPE, VOICING TYPE and GROUP on native Japanese speakers and learners' groups, subsets of the data were created which included only native Japanese speakers and learners of Japanese. A total of 10,038 responses from 42 participants across the two groups were fit into a generalised linear mixed effects model. Using stepwise regression for justifying the selected model, the best-fit model has a three-way interaction between VOICING TYPE, PRONUNCIATION TYPE, and GROUP. Trial number (zTrial) is removed as it does not show any effect. In addition, SUBJECT and WORD are added as random intercepts in the model. The results of the model is presented in Table 4.10. The reference level is voiced singleton in Japanese group and the predicted rating is 3.84.

```
lmer (response ~ PRONUNCIATION_TYPE * VOICING_TYPE * GROUP
+ (1+PRONUNCIATION_TYPE+VOICING_TYPE|SUBJECT)
+ (1+PRONUNCIATION_TYPE|WORD), data = Japanese and learners,
REML = F)
```

Table 4.10 Model summary for native Japanese speakers and learners of Japanese

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	3.84	0.11721	32.763	< 2e-16
VOICING_TYPE voiceless	-1.10667	0.07839	-14.117	< 2e-16
PRONUNCIATION_TYPE gemination	-0.65964	0.07499	-8.796	< 2e-16
PRONUNCIATION_TYPE wrong epenthesis	-2.05763	0.07496	-27.451	< 2e-16
PRONUNCIATION_TYPE CVC	-0.60144	0.07489	-8.031	1.08E-15
GROUP learners	-0.25506	0.16035	-1.591	0.117
VOICING_TYPE voiceless: PRONUNCIATION_TYPE gemination	2.08727	0.10596	19.698	< 2e-16
VOICING_TYPE voiceless: PRONUNCIATION_TYPE wrong epenthesis	0.87725	0.10603	8.274	< 2e-16
VOICING_TYPE voiceless: PRONUNCIATION_TYPE CVC	0.71017	0.10589	6.706	2.10E-11
VOICING_TYPE voiceless: GROUP learners	1.22705	0.10349	11.857	< 2e-16
PRONUNCIATION_TYPE gemination: GROUP learners	0.69662	0.1036	6.724	1.87E-11
PRONUNCIATION_TYPE wrong epenthesis: GROUP learners	0.61716	0.10358	5.958	2.63E-09

PRONUNCIATION_TYPE CVC: GROUP learners	-1.34514	0.10357	-12.987	< 2e-16
VOICING_TYPE voiceless: PRONUNCIATION_TYPE gemination: GROUP learners	-1.98108	0.14643	-13.53	< 2e-16
VOICING_TYPE voiceless: PRONUNCIATION_TYPE wrong epenthesis: GROUP learners	-1.13932	0.14653	-7.775	8.26E-15
VOICING_TYPE voiceless: PRONUNCIATION_TYPE CVC: GROUP learners	-0.70534	0.14646	-4.816	1.49E-06

Results for the interaction between VOICING TYPE, PRONUNCIATION TYPE, and GROUP is shown in Figure 4.9. The x-axes are the two groups and the y-axis is the predicted response rating. The light-blue solid lines refer to voiced stimuli and the pink dashed lines indicate voiceless stimuli.

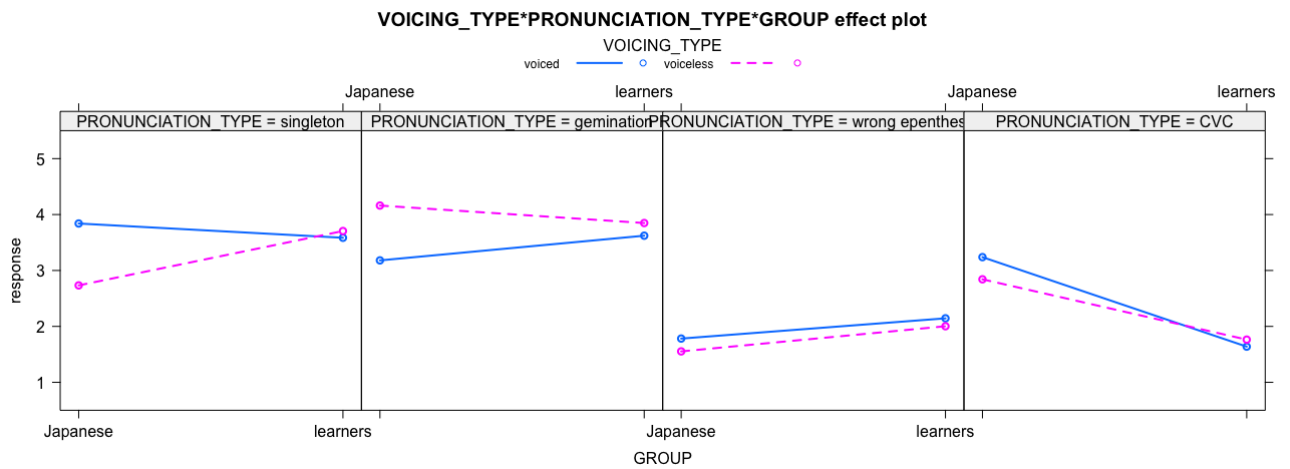


Figure 4.9 Plots of the three-way interaction between VOICING TYPE and PRONUNCIATION TYPE and GROUP.

It is a general pattern that L2 learners are less likely to be affected by the voicing type. That is, regardless of the voicing types in the coda position, their response patterns are very similar for each pronunciation type. In contrast, voicing type significantly affects the response patterns in native Japanese speakers for singleton and gemination. While Japanese participants rated voiced singletons higher than voiceless singletons, they rated voiceless geminates higher than voiced geminates. Thus, in the voiced context singletons are favoured, whereas in the voiceless context Japanese participants favoured geminations. Since it is predicted that rating for gemination would be influenced by place of articulation, this will be investigated further for each group later. Wrong epenthesis tends to receive lower ratings from both groups, suggesting participants acquire epenthetic rules. For CVC, while L2 learners confidently judged the stimuli as marginal Japanese pronunciation, native speakers of Japanese rated CVC higher. Results show that NZE-speaking learners of Japanese confidently find CVC would not be Japanese pronunciation more often than native Japanese speakers.

Each pronunciation type in relation to voicing types is summarised in Table 4.11. Shading marks indicate that statistically significant differences in ratings between voiceless and voiced contexts were observed, as predicted given prior literature. We expected different ratings for the two voicing contexts in singleton and gemination that reflect both the differences in probability distribution and L2 phonological grammar acquired from participants' lexicons.

As for Japanese participants, perceived well-formedness of singleton as well as gemination is related to voicing types of coda consonants in given stimuli. Regarding the effects of voicing, L2 learners show sensitivity to gemination but not to singleton. Important findings for wrong epenthesis in this section are that both native speakers of Japanese and L2 learners judged this pronunciation was not as good as singleton or gemination. The quality of epenthetic vowel is related to place of articulation. This will be investigated further for each group. There are significant differences between voiceless and voiced contexts in CVC for both groups, the effects of voicing were not directly relevant to CVC as well as the quality of epenthetic vowel.

Table 4.11 Comparison of predicted rating between voiceless and voiced contexts for each pronunciation type across Japanese group and learners' group. The effects of voicing were not directly relevant to the quality of epenthetic vowel as well as CVC.

	Japanese	Learners
Voiceless vs. Voiced:		
singleton	significant ( $p < 0.001$ )	not significant
gemination	significant ( $p < 0.001$ )	significant ( $p < 0.01$ )
wrong epenthesis	not significant	not significant
CVC	significant ( $p < 0.01$ )	significant ( $p < 0.01$ )

#### 4.4.3 Phonological process: Differences between learners and non-learners of Japanese

In order to examine whether the effects of pronunciation type and voicing type vary between learners and non-learners of Japanese, subsets of the data were created which included only these two groups and analysed. A total of 10,045 responses from 42 participants were fit into a generalised linear mixed effects model. As the same as the other models, stepwise regression was used for justifying the selected model. The best model to predict response patterns is below.

lmer (response ~ VOICING\_TYPE \* PRONUNCIATION\_TYPE +  
 PRONUNCIATION\_TYPE \* GROUP + zTrial + (1+zTrial|SUBJECT)  
 + (1|WORD), data = learners and non-learners, REML = F)

The best-fit model has the fixed effects of zTrial, two-way interactions between VOICING TYPE and PRONUNCIATION TYPE, and PRONUNCIATION TYPE and GROUP. In addition, SUBJECT and WORD are added as random intercepts, and random slopes of zTrial, by SUBJECT are included in the model. The results of the model are presented in Table 4.12. The reference level is voiced singleton in learners and the predicted rating is 3.60. There was a significant effect of zTrial (trial number) on the responses ( $p < 0.05$ ), with a negative estimate, indicating a significantly lower rating as the trial number increased. Plots of two-way interactions between pronunciation type and group, and pronunciation type and voicing type are shown in Figure 4.10 and Figure 4.11, respectively.

Table 4.12 Model summary for learners and non-learners of Japanese

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	3.60699	0.08657	41.668	< 2e-16
VOICING_TYPE voiceless	0.08672	0.0466	1.861	0.0634
PRONUNCIATION_TYPE gemination	0.05826	0.05194	1.122	0.26206
PRONUNCIATION_TYPE wrong epenthesis	-1.48338	0.05195	-28.554	< 2e-16
PRONUNCIATION_TYPE CVC	-1.94703	0.05199	-37.448	< 2e-16
GROUP non-learners	-0.27252	0.11983	-2.274	0.02715
zTrial	-0.04464	0.0198	-2.255	0.02941
VOICING_TYPE voiceless: PRONUNCIATION_TYPE gemination	0.06717	0.06088	1.103	0.26985
VOICING_TYPE voiceless: PRONUNCIATION_TYPE wrong epenthesis	-0.17502	0.06094	-2.872	0.00409
VOICING_TYPE voiceless: PRONUNCIATION_TYPE CVC	0.00899	0.06094	0.148	0.88272
PRONUNCIATION_TYPE gemination: GROUP non-learners	-0.11667	0.06092	-1.915	0.05551
PRONUNCIATION_TYPE wrong epenthesis: GROUP non-learners	1.15757	0.06098	18.981	< 2e-16
PRONUNCIATION_TYPE CVC: GROUP non-learners	1.15604	0.061	18.953	< 2e-16

In Figure 4.10, the x-axes are the two voicing types whereas the y-axes are the predicted response rating. The four lines refer to four pronunciation types in the two voicing context. The light-blue solid lines represent singleton, the pink dashed lines represent gemination, the green dotted lines represent wrong epenthesis, and the red dash-dotted lines represent CVC. Turning to the voicing effect on responses, although there is an interaction between pronunciation type



and voicing type, a difference was detected for gemination ( $p < 0.05$ ) and for CVC ( $p < 0.01$ ). In each voicing context, differences between singleton and gemination were not detected.

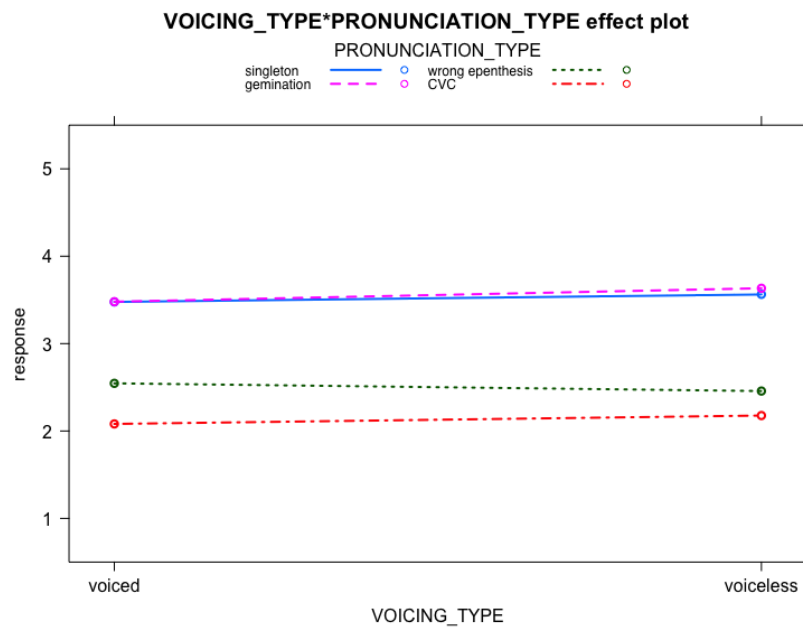


Figure 4.10 Plots of the two-way interaction between VOICING TYPE and PRONUNCIATION TYPE.

In Figure 4.11, the light-blue solid line refers to learners and the pink dashed line indicates non-learners. The interaction between pronunciation type and group is a result of different types of pronunciation being preferred for different groups. Since there is also a significant two-way interaction between voicing type and pronunciation type, investigating the results by pronunciation type is more informative.

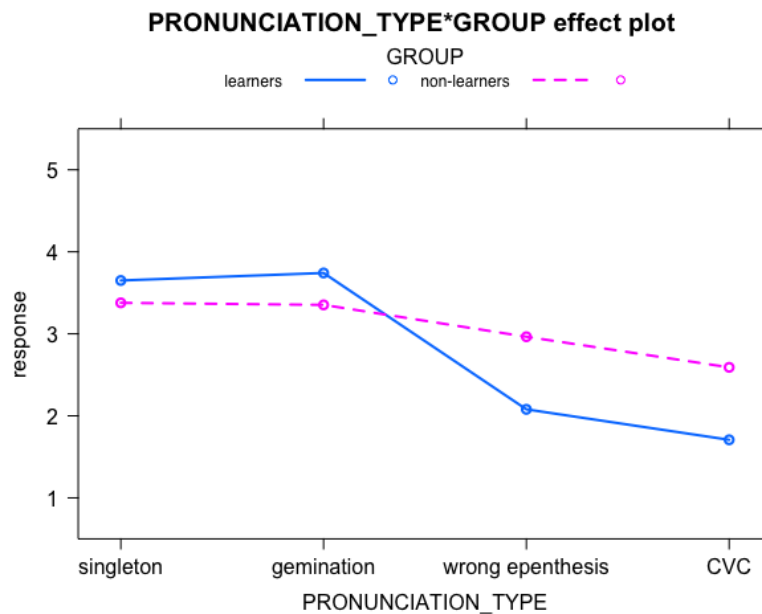


Figure 4.11 Plots of the two-way interaction between PRONUNCIATION TYPE and GROUP.

The effect of GROUP was found for singleton ( $p < 0.05$ ), epenthesis ( $p < 0.001$ ), and CVC ( $p < 0.001$ ), but not for gemination ( $p = 0.062$ ). The results suggest that learners of Japanese judge more confidently than non-learners for singleton which would be pronounced in Japanese. On the other hand, the lack of interaction for gemination and group indicates that their response patterns for these pronunciations are very similar for both groups. Learners of Japanese are more likely to judge confidently that wrong epenthesis would not be Japanese than non-learners. The ratings for CVC also showed the similar response patterns as wrong epenthesis.

To summarise, in general the effect of pronunciation type was found. In the full dataset, it seems that response patterns between learners and non-learners of Japanese were very similar at first glance, however, it was quite different for wrong epenthesis and CVC. Those stimuli are significantly disfavoured by learners of Japanese. The effect of voicing type upon these listeners was also found. However, importantly, it did not show the expected patterns in which participants would prefer singleton to gemination for voiced stimuli, and when gemination would be preferred for voiceless stimuli.

Next, the effects of other variables would be examined in a subset of the data for each group. For example, the effect of the place of articulation that shows a pronunciation type difference is explored separately in relation to each group. This was done by testing a VOICING TYPE x POA interaction with each respective group. In addition, metrics rating to an extra-linguistic factor (i.e., exposure to Japanese) would be explored.

#### 4.4.4 Phonological process: Native Japanese speakers

This section considers the results of native Japanese speakers, to establish a baseline of response patterns against which the predicted ratings can be compared. The 4,782 tokens from 20 participants were fitted in the generalised liner mixed effect model. There were three linguistic predictors and a trial number were initially included in this model. The three-way interaction model was conducted at the outset, and dimensions that were not significant were removed in a stepwise manner from the model. The best-fit model has a three-way interaction between VOICING TYPE, PRONUNCIATION TYPE, and POA. Along with these factors, SUBJECT is added as a random intercept in the model. The dependent variable was rating responses. The below is the best model predicting the response of the group. Trial number was not significant. Therefore, this factor was removed.

lmer (response ~ VOICING\_TYPE \* PRONUNCIATION\_TYPE \* POA +  
(1 |SUBJECT), data = Japanese, REML = F)

The results of the model are presented in Table 4.13. The reference category is the voiced labial singleton and the predicted rating was 4.39. There is a significant three-way interaction between voicing type, pronunciation type, and place of articulation. This indicates that different types of pronunciation are preferred for different voicing types and place of articulation.

Table 4.13 Model summary for native Japanese speakers

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	4.385	0.14440	30.361	< 2e-16
VOICING_TYPE voiceless	-1.73000	0.13630	-6.524	7.54E-11
PRONUNCIATION_TYPE gemination	-2.89900	0.13580	-12.69	< 2e-16
PRONUNCIATION_TYPE wrong epenthesis	-1.05500	0.13560	-21.347	< 2e-16
PRONUNCIATION_TYPE CVC	-0.88500	0.13560	-7.778	9.00E-15
POA alveolar	-1.48500	0.13560	-10.948	< 2e-16
POA velar	-0.15000	0.13560	-1.106	0.268862
VOICING_TYPE voiceless: PRONUNCIATION_TYPE gemination	2.23500	0.19230	11.622	< 2e-16
VOICING_TYPE voiceless: PRONUNCIATION_TYPE wrong epenthesis	0.64280	0.19210	3.347	0.000824
VOICING_TYPE voiceless: PRONUNCIATION_TYPE CVC	0.85360	0.19200	4.447	8.92E-06
VOICING_TYPE voiceless: POA alveolar	2.56000	0.19240	2.111	0.034806
VOICING_TYPE voiceless: POA velar	2.53800	0.19220	-5.578	2.57E-08
PRONUNCIATION_TYPE gemination: POA alveolar	0.90700	0.19200	13.3	< 2e-16
PRONUNCIATION_TYPE wrong epenthesis: POA alveolar	0.64520	0.19230	13.206	< 2e-16

PRONUNCIATION_TYPE CVC: POA alveolar	-0.00856	0.19210	4.725	2.37E-06
PRONUNCIATION_TYPE gemination: POA velar	0.45370	0.19200	3.355	0.000801
PRONUNCIATION_TYPE wrong epenthesis: POA velar	0.40500	0.19180	-0.045	0.964453
PRONUNCIATION_TYPE CVC: POA velar	-1.07000	0.19180	2.364	0.018126
VOICING_TYPE voiceless: PRONUNCIATION_TYPE gemination: POA alveolar	-1.65000	0.27170	-6.071	1.37E-09
VOICING_TYPE voiceless: PRONUNCIATION_TYPE wrong epenthesis: POA alveolar	-0.52830	0.27170	-1.944	0.05194
VOICING_TYPE voiceless: PRONUNCIATION_TYPE CVC: POA alveolar	-0.45560	0.27150	-1.678	0.093379
VOICING_TYPE voiceless: PRONUNCIATION_TYPE gemination: POA velar	1.21500	0.27170	4.472	7.94E-06
VOICING_TYPE voiceless: PRONUNCIATION_TYPE wrong epenthesis: POA velar	1.22400	0.27180	4.503	6.85E-06
VOICING_TYPE voiceless: PRONUNCIATION_TYPE CVC: POA velar	0.02769	0.27150	0.102	0.91877

Plots of three-way interactions between pronunciation type, voicing type, and place of articulation are shown in Figure 4.12. The x-axes are the two voicing types while the y-axis is the predicted response rating. The four lines refer to pronunciation types. The light-blue solid lines represent singleton, the pink dashed lines represent gemination, the green dotted lines represent wrong epenthesis, and the red dash-dotted lines represent CVC.

Overall, response patterns for labial (left panel plots) and velar (right panel plots) are similar, however that of alveolar (middle panel plots) is rather different. Remarkably, CVC receives high ratings, across all places of articulation regardless of voicing types. Wrong epenthesis tends to receive lower ratings than any other pronunciation types.

First, analysis of predicted response focuses on the relationship between voicing types and pronunciation types with place of articulation for singleton and gemination. The current study predicts that when C2 in given stimuli are voiced, singleton would receive higher ratings, but when C2 are voiceless consonants, gemination would receive higher ratings than singleton. These expected patterns can be seen for labial and velar in Figure 4.12. The figure clearly shows the effects of voicing type for labial and velar, as predicted. On the one hand, for voiced stimuli, singleton tends to receive higher ratings, yet on the other hand, gemination is more likely to receive higher ratings in the voiceless context. Although response patterns showed similarities between labial and velar, that of alveolar is different from these. For both voiced and voiceless contexts, gemination received higher ratings than singleton for alveolar. This means that gemination is preferred to singleton regardless of the voicing types. As with the full dataset, since interactions are detected and there are more than two levels in dependent variables, subsets are created to investigate further according to place of articulation.

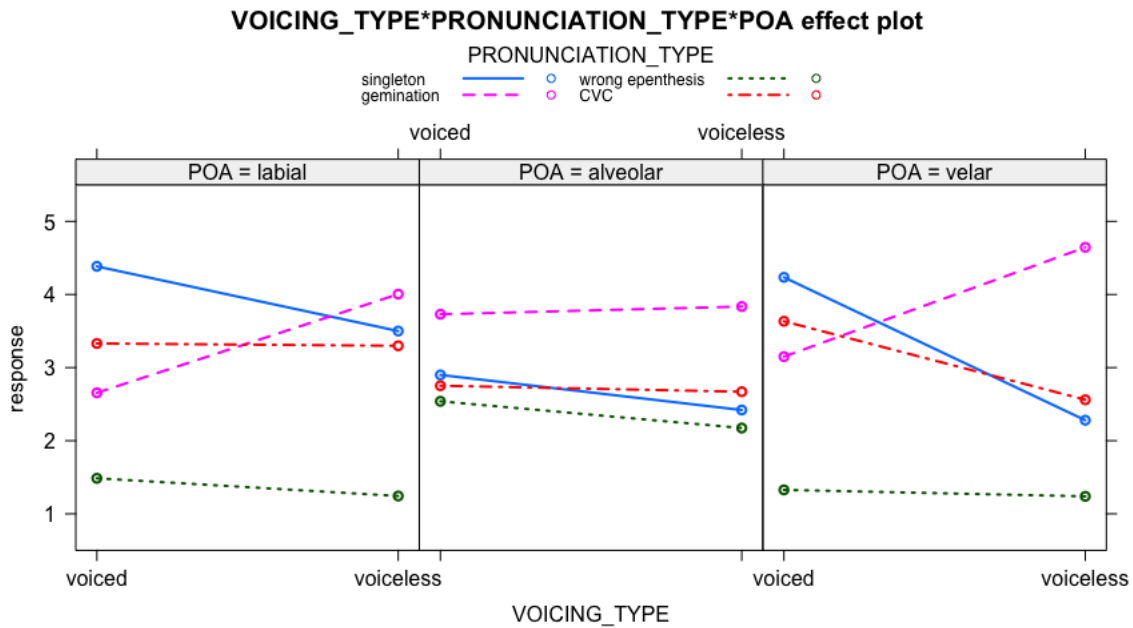


Figure 4.12 Plots of the three-way interaction between PRONUNCIATION TYPE, VOICING TYPE and POA.

The analysis of predicted response focuses on the effects of pronunciation type. As for the labial context, in the voiced context, singleton tends to receive higher ratings which are significantly different from CVC, gemination, and wrong epenthesis ( $p < 0.001$ ). Each pronunciation is significantly different from one another ( $p < 0.001$ ). For the voiceless context, gemination tends to receive higher ratings which is significantly different from singleton, CVC, and wrong epenthesis ( $p < 0.001$ ). There are no differences between singleton and CVC.

Similarly, for the velar, in the voiced context singleton tends to receive higher ratings which is significantly different from CVC, gemination, and wrong epenthesis ( $p < 0.001$ ). Each pronunciation is significantly different from each other ( $p < 0.001$ ). For the voiceless context, the predicted rating of gemination is significantly different from other pronunciation types, that is CVC, singleton, and wrong epenthesis ( $p < 0.001$ ), between CVC and singleton ( $p < 0.05$ ).

Lastly, as for the alveolar in the voiced context, gemination tends to receive higher ratings which is significantly different from singleton, CVC, and wrong epenthesis ( $p < 0.001$ ). There are no differences between singleton and CVC as well as between CVC and wrong epenthesis, respectively. However, wrong epenthesis is significantly different from singleton ( $p < 0.05$ ). For the voiceless context, gemination is also more likely to receive higher ratings than other pronunciation types, such as CVC, singleton and wrong epenthesis ( $p < 0.001$ ). There is a significant difference between CVC and wrong epenthesis ( $p < 0.05$ ), but no significant differences between singleton and CVC, or between singleton and wrong epenthesis, respectively.

Thus, results support the prediction that gemination tends to receive higher ratings with voiceless stimuli, whereas singleton is more likely to receive higher ratings in the voiced context for labial and velar places of articulation. Breaking down responses by place of articulation also shows gemination being significantly preferred for both voiced and voiceless alveolar stimuli to singleton. This is not surprising considering the frequency of voiced gemination when acknowledging the corpus-based loanword data: alveolar is the most frequent factor in gemination in loanwords. This will be further investigated below.

For wrong epenthesis, response patterns are also similar between labial and velar. Conversely, for alveolar it was rated higher than that of labial and velar contexts. This is an interesting result since CV sequences of wrong epenthesis in alveolars are illicit in the Japanese traditional CV inventory. The CV sequences /tu/ and /du/ are only used for loanwords, but examples of this are very rare (K. C. Hall, 2009). Investigating by subsets for the effect of place of articulation for wrong epenthesis confirms that alveolar is significantly more preferred than labial and velar in the voiced context ( $p < 0.001$ ) and in the voiceless context ( $p < 0.01$ ). There is no difference between labial and velar for both voicing types.

#### **4.4.4.1 Voiced geminates relation to POA**

Next, the effect of POA on voiced geminates is analysed. In the current study, place of articulation is particularly related to voiced geminates. POA is expected to influence the realisation of word-final stops in voiced contexts. Gemination in the voiced context can be expected in the following order: [dd] > [gg] > [bb]. Thus, participants with fine-grained knowledge will behave differently according to POA. In order to examine whether gemination occurs in this expected order, mean responses of individual subjects by place of articulation for each pronunciation type is calculated: that is, mean rated value for gemination minus that of non-gemination values. Figure 4.13 shows that gemination occurs in order of [dd] > [gg] > [bb] per native Japanese speaker. As expected, geminates occur more in the alveolar context than in other two contexts. Except for Subject 9 and 13, all participants rated alveolar higher than labial and velar. Regardless of the preferred pronunciation types, participants tend to respond in line with the prediction.

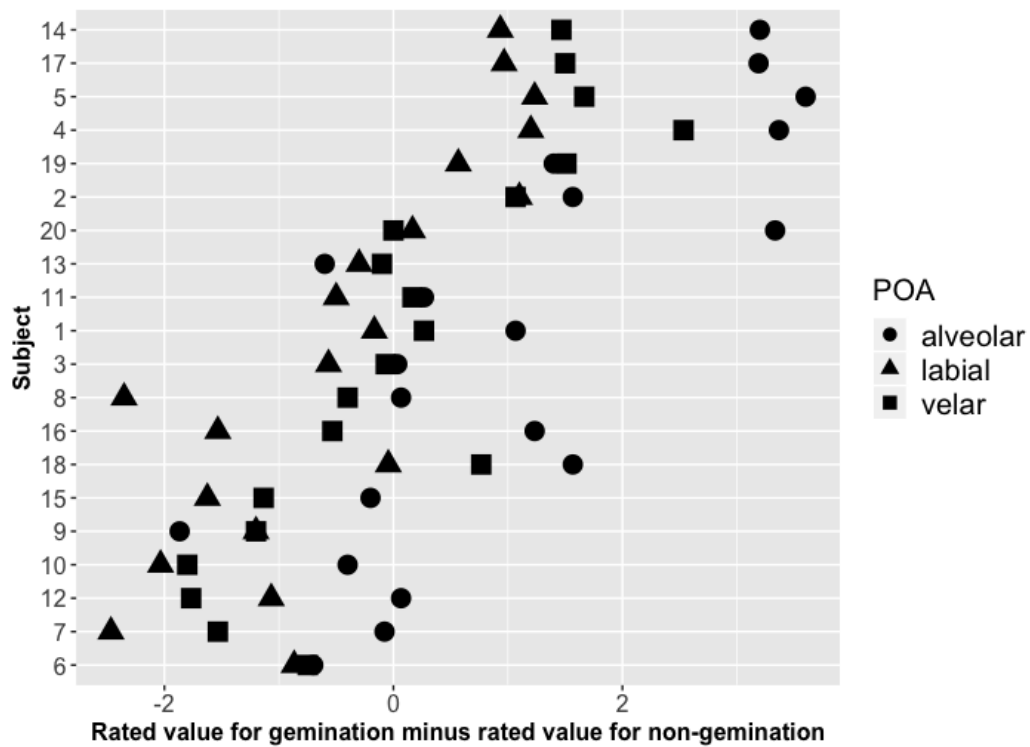


Figure 4.13 Mean response ratings of individual native Japanese speaker for gemination minus non-gemination ratings in voiced context by place of articulation.

#### 4.4.4.2 Rating for CVC

Although judgments on CVC is not our focus in this study, notably, results indicate that CVC received higher ratings than wrong epenthetic vowels, especially in the velar contexts. This pattern was observed in the pilot test as well. As mentioned previously, it is well known that native speakers of Japanese tend to perceive illusory vowels inside consonant clusters in stimuli (Dupoux et al., 1999; Dupoux et al., 2011). The higher ratings for CVC suggest that Japanese participants perceive a vowel after coda consonants, or the influence of English since all participants live in NZ. This would be investigated by using PCA scores: PC-English (exposure to English in NZ) and PC-Language (exposure to other languages in NZ). A post-hoc analysis was performed for a subset of CVC. In order to explore interactions between VOICING TYPE, POA, and PC-English and PC-Language, a separate model was run with two principal components revealed by PCA. The best predicted model is below.

$$\text{lmer}(\text{response} \sim \text{VOICING\_TYPE} * \text{POA} * \text{PC\_English} + \text{POA} * \text{PC\_Language} + (1 + |\text{SUBJECT}|), \text{data} = \text{CVC}, \text{REML} = \text{F})$$

Table 4.14 shows the model summary for the response pattern for the CVC pronunciation. The reference category is the voiced labial and the predicted ratings was 3.33. A three-way interaction between VOICING\_TYPE, POA and PC-English, and a two-way interaction between POA and PC-Language were detected, and this interaction is shown in Figure 4.14 and Figure 4.15. In Figure 4.14, the x-axes present individuals' PC-English scores while the y-axis is the predicted response ratings. A higher number orientates to more exposure to English in NZ. The light-blue solid lines refer to voiced stimuli in which C2 is voiced stops, and the pink dashed lines indicate voiceless stimuli with voiceless stops in C2.

Table 4.14 Model summary for CVC

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	3.33	0.21575	15.434	1.61E-14
VOICING_TYPE voiceless	-0.02648	0.1152	-0.23	0.81826
POA alveolar	-0.57258	0.1152	-4.97	7.68E-07
POA velar	0.30137	0.1152	2.616	0.00901
PC-ENGLISH	-0.65002	0.22966	-2.83	0.00895
PC-LANGUAGE	-0.19932	0.22194	-0.898	0.37878
VOICING_TYPE voiceless: POA alveolar	-0.06094	0.16292	-0.374	0.70843
VOICING_TYPE voiceless: POA velar	-1.04489	0.16292	-6.414	2.05E-10
VOICING_TYPE voiceless: PC_English	-0.17534	0.11837	-1.481	0.13879
POA alveolar: PC_English	-0.03557	0.12066	-0.295	0.76822
POA velar: PC_English	-0.05681	0.12046	-0.472	0.63727
POA alveolar: PC-Language	-0.2612	0.08684	-3.008	0.00269
POA velar: PC-Language	0.16371	0.08691	1.884	0.05986
VOICING_TYPE voiceless: POA alveolar: PC_English	0.28325	0.16733	1.693	0.09076
VOICING_TYPE voiceless: POA velar: PC-English	0.51276	0.16717	3.067	0.00221



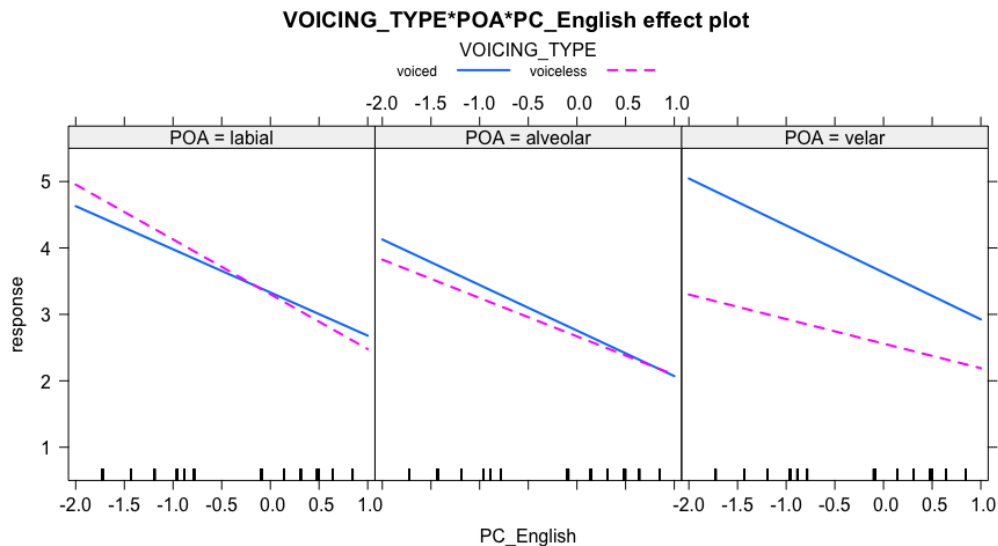


Figure 4.14 Plots of the three-way interaction between VOICING TYPE, POA, and PC-ENGLISH (exposure to English in NZ) in the native speakers of Japanese dataset.

The plots in Figure 4.14 show that the response pattern of participants with less exposure to English are more affected by voicing types than participants who have more exposure to English. More importantly, the effect of PC-English indicates that participants with more exposure to English in NZ are more likely to give lower ratings to CVC than participants with less exposure to English. This might indicate that the more people learn English the less likely they are to perceive illusory vowels after final consonants.

There is also an interaction between POA and PC-Language (exposure to other languages in NZ) as shown in Figure 4.15. The x-axis presents individuals' PC-Language scores and the y-axis is the predicted response ratings. A higher number orientates to more exposure to other languages in NZ. The three-line types in the plot refer to the three places of articulation. The light-blue solid line represents labial, the pink dashed line represents alveolar, and the green dotted lines represents velar. The figure shows that participants who have more exposure to other languages are more influenced by place of articulation. They are more likely to judge alveolar CVC stimuli as an illicit pronunciation. The results of the three-way interaction and two-way interaction suggest that even for participants with higher PC-English, CVC for velar was not easy to judge.

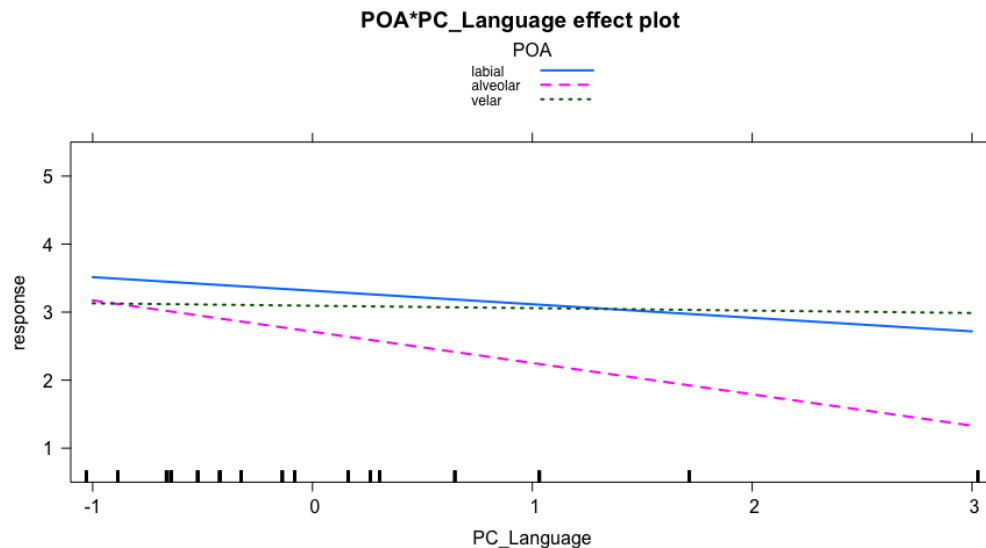


Figure 4.15 Plots of the two-way interaction between POA and PC-Language (exposure to other languages).

To summarise, the results showed that pronunciation type is a significant predictor of well-formedness of loanword pronunciation in Japanese. When word-final consonants are voiced, singletons were preferred, whereas in cases of voiceless stimuli, geminates were judged well-formed. The interaction between pronunciation type, voicing type, and place of articulation suggests that the response patterns of native speakers of Japanese are relatively consistent with the observed patterns in nativised loanwords, suggesting that native Japanese speakers have knowledge of loanword phonology in Japanese. Extra-linguistic factors also play a role in the judgement of CVC stimuli, in that participants with higher exposure to English are more likely to reject the pronunciation.

#### 4.4.5 Phonological process: Learners of Japanese

This section considers the results of NZE-speaking learners of Japanese. The 5,256 tokens from 22 participants were fitted to a generalised liner mixed effect model. There were three linguistic predictors and two extra-linguistic predictors, and all factors were included in the model initially. The four-way interaction model was conducted at the outset, and dimensions that were not significant were removed in a stepwise manner from the model. The best-fit model has the fixed effects of PC-JJ (exposure to Japanese in Japan) and zTrial, a three-way interaction between VOICING TYPE, PRONUNCIATION TYPE and POA, and a three-way-interaction between VOICING TYPE, PRONUNCIATION TYPE and PC-JNZ (exposure to Japanese in NZ). Along with these factors, SUBJECT is added as random intercept, and a random slope of zTrial by SUBJECT is included in the model.

lmer (response ~ VOICING\_TYPE \* PRONUNCIATION\_TYPE \* POA +  
 VOICING\_TYPE \* PRONUNCIATION\_TYPE \* PC-JNZ + PC-JJ +  
 zTrial +(1+zTrial|SUBJECT), data = learners, REML = F)

As can be seen in the model above, there is lack of interaction between pronunciation type, voicing types, and PC-JJ (exposure to Japanese in Japan). This indicates that response patterns between voiced and voiceless stimuli for pronunciation types would be very similar regardless of whether participants had exposure to Japanese in Japan or not.

The results of the model are presented in Table 4.15. The reference category is the labial singleton, and the predicted rating was 3.75. Results showed the main effects of PC-JJ and zTrial. While PC-JJ appears to have a positive significant effect on the response ( $p < 0.05$ ), the number of trials appear to have a negative significant effect on the responses ( $p < 0.05$ ). The effect of PC-JJ indicates that participants with higher exposure to Japanese in Japan tend to give a rating of 0.24 or higher for stimuli in general, regardless of pronunciation types. The effect of trials gives a rating of -0.063, which was lower over the course of the experiment. There was also a significant three-way interaction between pronunciation type, voicing type, and place of articulation. The other significant three-way interaction was pronunciation type, voicing type, and PC-JNZ (exposure to Japanese in NZ).

Table 4.15 Model summary for learners of Japanese

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	3.75168	0.1092	34.357	< 2e-16
VOICING_TYPE voiceless	0.11035	0.10756	-0.515	0.606699
PRONUNCIATION_TYPE gemination	-0.0554	0.10761	-17.734	< 2e-16
PRONUNCIATION_TYPE wrong epenthesis	-1.907	0.10753	-19.219	< 2e-16
PRONUNCIATION_TYPE CVC	-2.07051	0.10773	1.026	0.304984
POA alveolar	-0.56495	0.10752	-5.255	1.54E-07
POA velar	0.06356	0.10739	0.592	0.553969
PC-JNZ	-0.04527	0.09679	-0.468	0.643042
PC-JJ	0.24001	0.08722	2.752	0.011641
zTrial	-0.06305	0.02703	-2.332	0.029206
VOICING_TYPE voiceless: PRONUNCIATION_TYPE gemination	0.02983	0.15208	0.196	0.844529
VOICING_TYPE voiceless: PRONUNCIATION_TYPE wrong epenthesis	-0.01905	0.15215	-0.125	0.900387
VOICING_TYPE voiceless: PRONUNCIATION_TYPE CVC	0.12452	0.15221	0.818	0.413341
VOICING_TYPE voiceless: POA alveolar	-0.07318	0.15216	2.063	0.039157
VOICING_TYPE voiceless: POA velar	0.09748	0.15187	8.541	< 2e-16

PRONUNCIATION_TYPE gemination: POA alveolar	0.31377	0.15209	2.954	0.003146
PRONUNCIATION_TYPE wrong epenthesis: POA alveolar	1.29873	0.15206	-0.271	0.786061
PRONUNCIATION_TYPE CVC: POA alveolar	0.44989	0.15227	0.683	0.494542
PRONUNCIATION_TYPE gemination: POA velar	-0.04125	0.15196	-0.538	0.590595
PRONUNCIATION_TYPE wrong epenthesis: POA velar	0.10379	0.15193	-0.481	0.6306
PRONUNCIATION_TYPE CVC: POA velar	-0.08185	0.15214	0.642	0.520996
VOICING_TYPE voiceless: PC-JNZ	0.21149	0.06344	3.807	0.000142
PRONUNCIATION_TYPE gemination: PC-JNZ	0.24164	0.06346	2.55	0.010794
PRONUNCIATION_TYPE wrong epenthesis: PC-JNZ	0.16199	0.06352	6.128	9.53E-10
PRONUNCIATION_TYPE CVC: PC-JNZ	0.38883	0.06345	3.334	0.000863
VOICING_TYPE voiceless: PRONUNCIATION_TYPE gemination: POA alveolar	0.05092	0.21517	0.237	0.812953
VOICING_TYPE voiceless: PRONUNCIATION_TYPE wrong epenthesis: POA alveolar	-0.3585	0.21537	-1.665	0.096052
VOICING_TYPE voiceless: PRONUNCIATION_TYPE CVC: POA alveolar	-0.08088	0.21528	-0.376	0.707157
VOICING_TYPE voiceless: PRONUNCIATION_TYPE gemination: POA velar	0.19798	0.21473	0.922	0.356577
VOICING_TYPE voiceless: PRONUNCIATION_TYPE wrong epenthesis: POA velar	-0.36507	0.21477	-1.7	0.089233
VOICING_TYPE voiceless: PRONUNCIATION_TYPE CVC: POA velar	-0.26291	0.21501	-1.223	0.221459
VOICING_TYPE voiceless: PRONUNCIATION_TYPE gemination: PC-JNZ	-0.09386	0.08969	-1.046	0.29541
VOICING_TYPE voiceless: PRONUNCIATION_TYPE wrong epenthesis: PC-JNZ	-0.188	0.08981	-2.093	0.036362
VOICING_TYPE voiceless: PRONUNCIATION_TYPE CVC: PC-JNZ	-0.16663	0.08968	-1.858	0.063216

#### 4.4.5.1 Effects of phonological factors

First, the analysis of predicted response focuses on the relationship between voicing type, pronunciation types, and place of articulation as shown in Figure 4.16. The x-axes are the two voicing types and the y-axis is the predicted response ratings. The four lines refer to each pronunciation type. The light-blue solid lines represent singleton, the pink dashed lines represent gemination, the green dotted lines represent wrong epenthesis, and the red dash-dotted lines represent CVC.

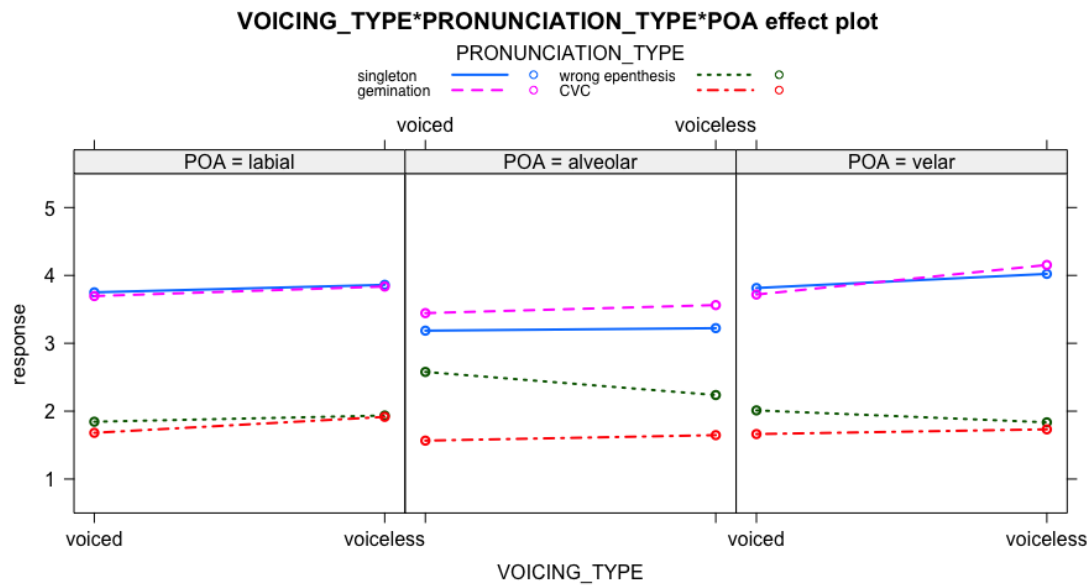


Figure 4.16 Plots of the three-way interaction between VOICING TYPE, PRONUNCIATION TYPE and POA in the learners of Japanese dataset.

This model is related to the research questions stated, regarding **RQ1** whether a sublexicon phonology of a language can be learnt by exposure to the target language, as it was predicted that the learning of phonological rules is possible without being taught, and regarding **RQ2**, if any, what rules are implicitly learned? Overall, response patterns are very similar between labial (left panel plots) and velar (right panel plots). The two panels show that, regardless of voicing types, singleton and gemination tend to receive higher ratings whereas wrong epenthesis and CVC are more likely to receive low ratings. On the other hand, for alveolar contexts (middle panel), wrong epenthesis did not receive lower ratings in comparison to the other two contexts. To explore these patterns further, the data were further split into subsets by place of articulation, and the effects of the two-way interactions between voicing types and pronunciation type were examined separately.

For the labial context, there are no differences between singleton and gemination in each voicing type. Also, there are no differences for response patterns between wrong epenthesis and CVC in each voicing type. Response patterns of singleton and gemination are significantly different from those of wrong epenthesis and CVC ( $p < 0.001$ ). For the velar in the voiced context, although singleton received slightly higher ratings than gemination, there are no differences between singleton and gemination. However, there is a significant difference in response patterns between wrong epenthesis and CVC ( $p < 0.01$ ). In the voiceless context, there are no differences between gemination and singleton as well as between wrong epenthesis and CVC. Response patterns of singleton and gemination are significantly different from those

of wrong epenthesis and CVC ( $p < 0.001$ ). Lastly, for the alveolar context, the effects of pronunciations were found between each pronunciation type in both voicing contexts: in the voiced context, gemination  $>$  ( $p < 0.05$ ) singleton  $>$  ( $p < 0.001$ ) wrong epenthesis  $>$  ( $p < 0.001$ ) CVC; in the voiceless context, gemination  $>$  ( $p < 0.01$ ) singleton  $>$  ( $p < 0.001$ ) wrong epenthesis  $>$  ( $p < 0.001$ ) CVC. Thus, only in this context, the response patterns in this context exhibit the observed nativised patterns in loanwords. Gemination was significantly preferred for voiced stimuli to singleton. This pattern was observed in Japanese group as well.

The results suggest that the effect of pronunciation is similar for labial and velar, but slightly different for alveolar. The results are summarised in Table 4.16. The effects of pronunciation type suggest that L2 learners of Japanese know well-formed Japanese pronunciation. Especially in relation to labial and velar contexts, wrong epenthesis and CVC were judged as falling short of clear Japanese pronunciation. This means that the participants have a good knowledge of Japanese phonology in general, in that final consonants in word-final position are disallowed and must be repaired through appropriate epenthetic vowels. As with the results of the native speakers of Japanese, the response patterns for alveolar contexts are slightly differently. This will be discussed below.

Table 4.16 Summary table for effects of pronunciation type on responses

Labial	Voiced	Singleton, Gemination $>$ ( $p < 0.001$ ) Wrong epenthesis, CVC
	Voiceless	Singleton, Gemination $>$ ( $p < 0.001$ ) Wrong epenthesis, CVC
Alveolar	Voiced	Gemination $>$ ( $p < 0.05$ ) $>$ Singleton $>$ ( $p < 0.001$ ) Wrong epenthesis $>$ ( $p < 0.001$ ) CVC
	Voiceless	Gemination $>$ ( $p < 0.01$ ) $>$ Singleton $>$ ( $p < 0.001$ ) Wrong epenthesis $>$ ( $p < 0.001$ ) CVC
Velar	Voiced	Singleton, Gemination $>$ ( $p < 0.001$ ) Wrong epenthesis $>$ ( $p < 0.01$ ) CVC
	Voiceless	Gemination, Singleton $>$ ( $p < 0.001$ ) Wrong epenthesis, CVC

This study predicts that when word-final stops are voiceless, listeners with knowledge of Japanese prefer gemination. On the other hand, singletons would be preferred when word-final stops are voiced. As far as the effect of pronunciation is concerned, the predictions were not borne out since the rating pattern between singleton and gemination does not show any significant differences in the labial and velar contexts. These results indicate that the participants do not have knowledge that final voiceless stops preceded by a lax vowel geminate almost all the time in nativised loanwords in Japanese.

Finally, in order to explore the effect of the place of articulation, the data were subset by pronunciation type. The current study predicts that gemination in the voiced context can be expected in the following order: [dd]  $>$  [gg]  $>$  [bb]. This hierarchical model is based on

gemination rates in nativised loanwords in Japanese. Contrary to the prediction, for gemination in the voiced context, velar and labial received higher ratings than alveolar ( $p < 0.01$ ). The difference was not found between velar and labial.

Thus, as predicted, L2 learners of Japanese have acquired epenthetic vowels (i.e., Rule A) rather than stochastic patterns of consonant gemination (i.e., Rule B). Thus, categorical rules have been more readily learned than gradient ones. Regarding **RQ 3**, are L2 learners with high exposure to Japanese able to acquire fine-grained knowledge regarding the effects of voicing and place of articulation (POA)? Since POA did not interact with PC effects, L2 learners were not influenced by voicing types and place of articulation in relation to their levels of exposure to Japanese.

#### 4.4.5.2 Effects of extra-linguistic factors

Next, the analysis of predicted response focuses on the relationship between pronunciation type, voicing type, and the extra-linguistics factor PC-JNZ (exposure to Japanese in NZ). The results are shown in Figure 4.17. The x-axes present individuals' PC-JNZ scores. A higher PC-JNZ indicates that participants have more exposure to Japanese in NZ than participants with lower PC-JNZ scores. The y-axis is the predicted response rating. The four lines refer to each pronunciation type. The light-blue solid lines represent singleton, the pink dashed lines represent gemination, the green dotted lines represent wrong epenthesis, and the red dash-dotted lines represent CVC. The effect of PC-JNZ predicts that participants with higher PC-JNZ scores are more likely to be sensitive to voicing types for gemination and singletons.

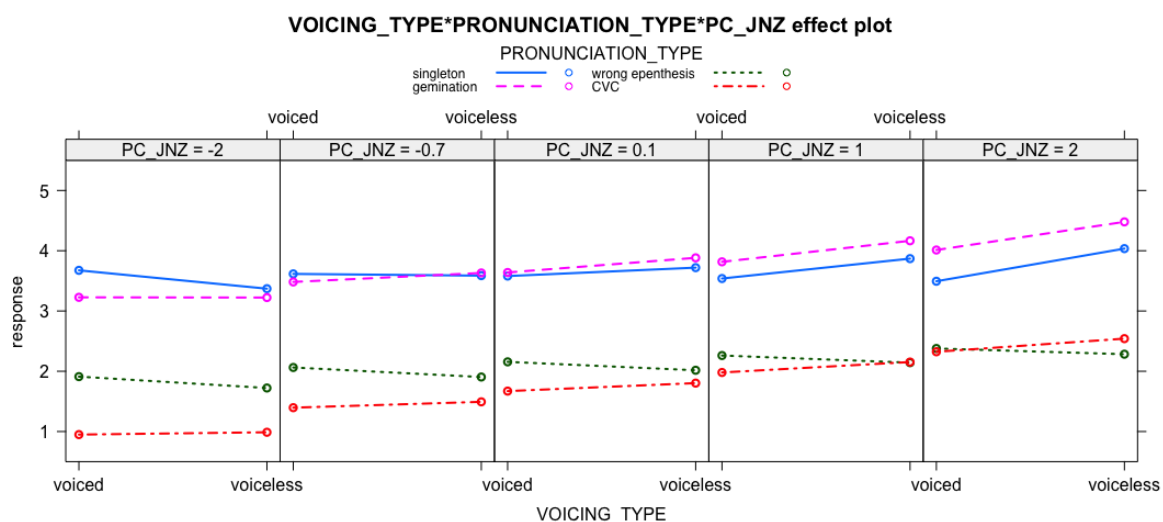


Figure 4.17 Plots of the three-way interaction between VOICING TYPE, PRONUNCIATION TYPE and PC-JNZ in the learners of Japanese dataset.

Figure 4.17 shows that voicing types tend to influence listeners' response patterns for singleton and gemination in relation to PC-JNZ. However, the effect of voicing type for singleton was opposite to what was expected in the current study. Participants with less exposure to Japanese are more likely to rate singleton higher when C2 in the stimuli are voiced. This pattern would be expected for participants with higher PC-JNZ scores. However, as the score keeps increasing, gemination received higher ratings than singleton in the same voiced context. In the voiceless context, participants with less exposure to Japanese are not influenced by voicing types for singleton and gemination, while participants with more exposure to Japanese are likely to judge voiceless stimuli are better for gemination. There are no significant interactions between voicing and PC-JNZ for other pronunciations. For CVC, participants with more exposure to Japanese seem less confident in judging CVC as illicit pronunciation than participants with less exposure.

Regarding **RQ 3**, are L2 learners with high exposure to Japanese able to acquire fine-grained knowledge regarding the effects of voicing and place of articulation (POA)? Since POA did not interact with PC effects, L2 learners were not influenced by voicing types and place of articulation in relation to their levels of exposure to Japanese.

#### **4.4.6 Phonological process: Non-learners of Japanese**

Finally, the results of NZE-speaking non-learners of Japanese are considered. The 4,789 tokens from 20 participants were fitted in the generalised liner mixed effect model. There are three linguistic predictors and one extra-linguistic predictor, and all factors were initially included in the model. The four-way interaction model was fitted, and factors that are not significant were removed in a stepwise manner from the model. Thus, the best fit model does not include VOICING\_TYPE as a predictor. There was a lack of interaction for both pronunciation types and place of articulation with voicing types. This indicates that response patterns are very similar regardless of whether voicing type is voiced or voiceless in C2 for each pronunciation type and for each place of articulation. The best-fit model had a two-way interaction between PRONUNCIATION TYPE and POA, and PRONUNCIATION TYPE and PC-JNZ with SUBJECT as a random intercept.

```
lmer (response ~ PRONUNCIATION_TYPE * POA +
PRONUNCIATION_TYPE * PC-JNZ + (1|SUBJECT),
data = non-learners, REML = F)
```



The results of the model are presented in Table 4.17. The reference category is the labial singleton and the predicted rating is 3.43. Results show that there is a two-way interaction between PRONUNCIATION TYPE and POA and interaction between PRONUNCIATION TYPE and PC-JNZ. The two-way interactions between pronunciation type and place of articulation, and pronunciation type and PC-JNZ are shown in Figure 4.18 and Figure 4.19 respectively.

Table 4.17 Model summary for non-learners of Japanese

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	3.42600	0.07084	-0.053	0.9575
PRONUNCIATION_TYPE gemination	-0.00370	0.06940	-7.859	4.75E-15
PRONUNCIATION_TYPE wrong epenthesis	-0.54570	0.06944	-11.335	< 2e-16
PRONUNCIATION_TYPE CVC	-0.78710	0.06944	-4.196	2.76E-05
POA alveolar	-0.29120	0.06940	2.396	0.0166
POA velar	0.16630	0.06940	2.131	0.0415
PC-JNZ	0.12760	0.05987	-0.268	0.7886
PRONUNCIATION_TYPE gemination: POA alveolar	-0.02630	0.09808	6.311	3.02E-10
PRONUNCIATION_TYPE wrong epenthesis: POA alveolar	0.61940	0.09814	2.085	0.0371
PRONUNCIATION_TYPE CVC: POA alveolar	0.20460	0.09811	-0.472	0.6369
PRONUNCIATION_TYPE gemination: POA velar	-0.04630	0.09808	-2.23	0.0258
PRONUNCIATION_TYPE wrong epenthesis: POA velar	-0.21900	0.09820	-1.996	0.046
PRONUNCIATION_TYPE CVC: POA velar	-0.19600	0.09820	2.102	0.0356
PRONUNCIATION_TYPE gemination: PC-JNZ	0.08631	0.04106	0.547	0.5845
PRONUNCIATION_TYPE wrong epenthesis: PC-JNZ	0.02246	0.04107	-4.661	3.23E-06
PRONUNCIATION_TYPE CVC: PC-JNZ	-0.19150	0.04108	-0.053	0.9575

#### 4.4.6.1 Effects of phonological factors

First, the analysis of predicted response focuses on the relationship between pronunciation types and place of articulation, as shown in Figure 4.18. The x-axis presents the four pronunciation types and the y-axis presents the predicted response rating. The light-blue solid line refers to the labial context, the pink dashed line indicates the alveolar context, and the green dotted line is the velar context. The results show that singleton and gemination in labial and velar contexts are likely to receive a higher rating than that of alveolars. CVC tends to receive a lower rating across all places of articulation. This indicates that NZE-speaking non-learners of Japanese disfavoured CVC for all three of place of articulation. Wrong epenthesis in labial and velar contexts are judged in a similar manner. The results suggest that non-learners know more than Japanese CV syllable structure. If the participants only know that closed

syllables in word-final position are prohibited except with a moraic nasal in Japanese, only CVC would receive lower rating. When the pronunciation type was wrong epenthesis, the response rating for labial and velar decreased, whereas that of alveolar increased. This will be examined further in the Discussion (§4.5.2)

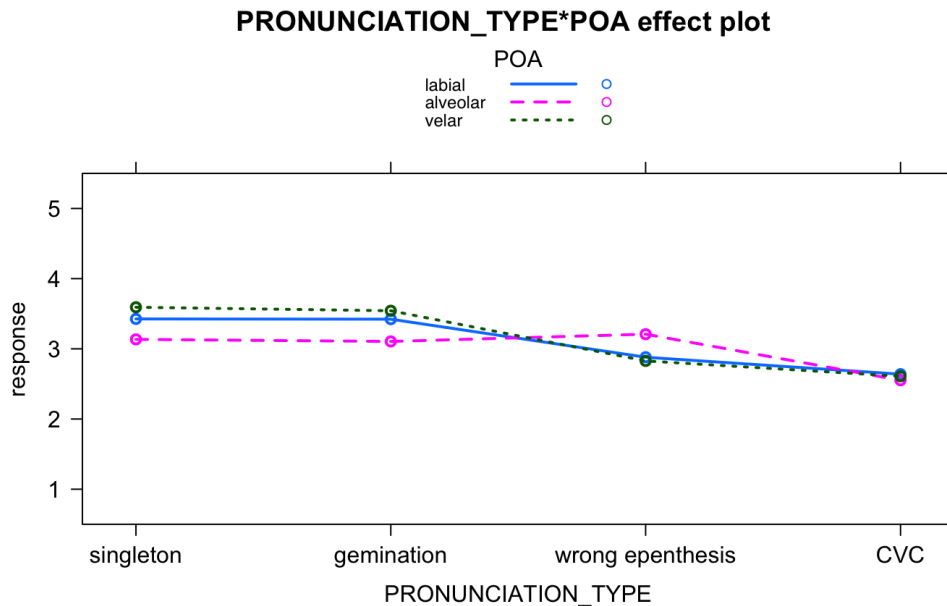


Figure 4.18 Plots of the two-way interaction between PRONUNCIATION TYPE and POA for non-learners of Japanese.

Because of the presence of two-way interaction between pronunciation type and place of articulation, a separate multiple comparison analysis for the effect of pronunciation for each place of articulation was conducted by using subsets. Breaking down responses by place of articulation shows similar patterns for labial and velar contexts across pronunciation type: there is more preference for gemination and singleton that are licit pronunciation types in Japanese than CVC and wrong epenthesis. On the other hand, alveolar shows a different pattern from labial and velar contexts: wrong epenthesis is more likely to be favoured than other two contexts, as shown in Table 4.18.

Table 4.18 Summary comparison of predicted rating for each place of articulation (Intercept = singleton)

Labial	Singleton, Gemination > ( $p < 0.001$ ) Wrong Epenthesis > ( $p < 0.001$ ) CVC
Alveolar	Wrong Epenthesis, Singleton, Gemination > ( $p < 0.001$ ), CVC
Velar	Singleton, Gemination > ( $p < 0.001$ ) Wrong Epenthesis ( $p < 0.01$ ) CVC

For labial contexts, no significant difference was found between singleton and gemination responses. However, singleton and gemination results were significantly different to wrong epenthesis and CVC ( $p < 0.001$ ). The difference between wrong epenthesis and CVC was also significant ( $p < 0.001$ ). These patterns are also found for velar. For alveolar, there are no significant differences between wrong epenthesis, singleton, and gemination. Wrong epenthesis and singleton are significantly different from CVC ( $p < 0.001$ ). In addition, there is the effect of the place of articulation in which wrong epenthesis for labial and velar is more significantly disfavoured than for alveolar ( $p < 0.001$ ). For singleton and gemination, labial and velar tend to receive significantly higher ratings than alveolar ( $p < 0.001$ ). For CVC, the effect of place was not found.

#### **4.4.6.2 Effects of extra-linguistic factors**

Next, the analysis of predicted responses focuses on the relationship between pronunciation types and the extra-linguistic factor PC-JNZ (exposure to Japanese in NZ). Figure 4.19 illustrates the interaction PRONUNCIATION TYPE and PC-JNZ. The x-axis presents individuals' PC-JNZ scores and the y-axis is the predicted response ratings. The four-line types in the plot refer to the four pronunciation types. The light-blue solid line refers singleton, the pink dashed line indicates gemination, the green dotted lines represent wrong epenthesis, and the red dash-dotted lines represent CVC. A higher PC-JNZ indicates that participants have more exposure to Japanese in NZ than participants with lower PC-JNZ scores. The effect of PC-JNZ predicts that participants with higher PC-JNZ scores are more likely to give licit pronunciation in Japanese a higher rating. Conversely, a negative correlation is expected between the response and PC-JNZ for illicit pronunciation types.

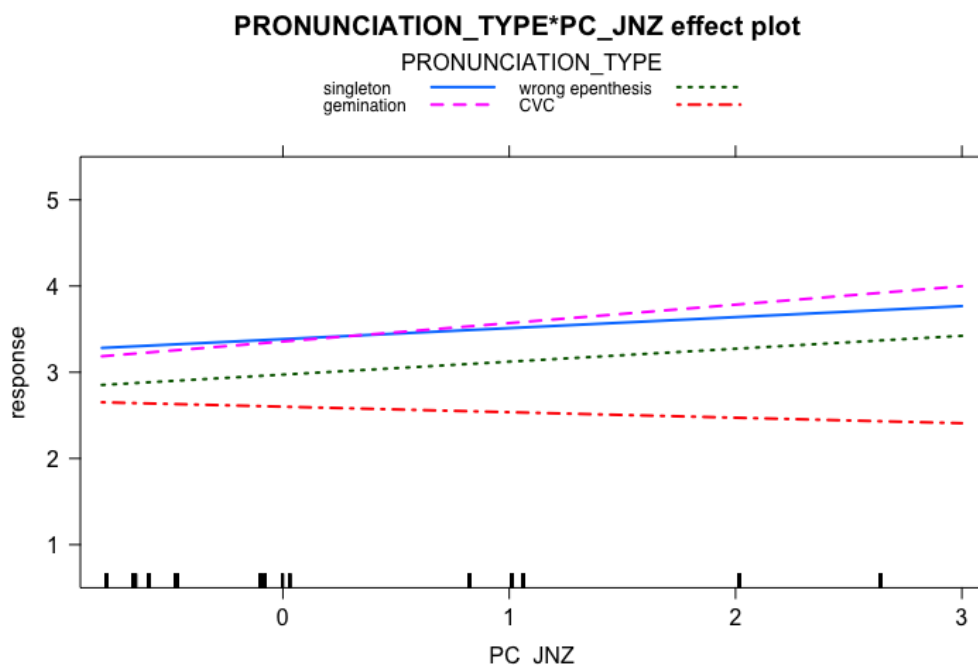


Figure 4.19 The plot of the interaction between PRONUNCIATION TYPE and PC-JNZ (exposure to Japanese in NZ).

There is a significant interaction between PRONUNCIATION TYPE and PC-JNZ due to differences between singleton and gemination ( $p < 0.05$ ), as well as between singleton and CVC ( $p < 0.001$ ). Participants with more exposure to Japanese are more likely to rate gemination higher. More importantly, the plot shows that the response patterns of participants with less exposure to Japanese are not affected by pronunciation types. On the other hand, participants who have more exposure to Japanese are less likely to give each pronunciation type the same rating. For CVC only, there is a negative relation between response and PC-JNZ. The exposure to Japanese is likely to play a role in judging well-formedness of the language without being taught. Thus, the findings on the effect of the extra-linguistic factor are relatively consistent with the prediction of the current study, contrary to the group of learners.

## 4.5 Discussion

The study presented in this chapter investigated the extent to which native and non-native speakers of Japanese learn loanword phonology through experience of using and/or passive exposure to that language. In addition to assessing speakers' knowledge of L2 loanword phonological regularities in Japanese, the study examined the influence of phonotactic patterns in the Japanese lexicon.

This study asked the following questions: **RQ1** is it possible that a sublexicon phonology of a language is learned from exposure to the target language? **RQ2**: If any, what rules are implicitly learned in relation to English word-final stop consonants? (i.e., epenthetic rule or geminates or both) **RQ3**: Are L2 learners with high exposure to Japanese able to acquire fine-grained knowledge regarding the effects of voicing and place of articulation (POA)? **RQ4**: Does the language's overall statistical patterns influence learners' response patterns? Specifically, would participants be biased in responding with the most expected pattern in the language rather than with observed patterns in the Japanese loanwords?

Findings reveal that both native Japanese speakers and L2 learners of Japanese accessed loanword phonotactics during the auditory well-formedness experiment rather than overall Japanese phonotactics. That is, patterns of ratings are congruent with logged scores of sublexicon phonotactics rather than that of Japanese phonotactics as a whole. These results reflect a general cognitive ability to detect not only the statistical properties of a language but also sublexicon phonology in the language. As a Bayesian learning-based computational clustering model by Morita (2018) predicted, differences in probability distribution enable language users to detect the property of sublexicons in Japanese. The results of NZE-speaking learners of Japanese show similarities and differences in response patterns from native speakers of Japanese. Although L2 learners' performance has not achieved the level of native Japanese speakers, they have some knowledge of the adaptations needed for loanwords in Japanese. This section begins with a summary of findings, and then presents the discussion of the acquisition of a sublexicon phonology in the language with findings related specifically to L2 learners of Japanese.

#### **4.5.1 Summary of results**

##### **4.5.1.1 Categorical rules: Quality of epenthetic vowels**

The results for the perceptual well-formedness experiment for the pronunciation of loanwords demonstrate that native speakers of Japanese have knowledge of observed patterns of nativised loanwords in Japanese. This is expected from research that has looked into statistical learning in first languages (e.g., Edwards et al., 2004; Vitevitch & Luce, 1998; Zamuner et al., 2004) and studies of Japanese stratum-specific phonotactics (e.g., Gelbart & Kawahara, 2007; Moreton, 2002; Moreton & Amano, 1999; Morita, 2018). Native Japanese speakers' response patterns are relatively consistent with the observed patterns in loanwords, except for the ratings of CVC where misperception of the stimuli may have occurred.

As for NZE-speaking learners of Japanese, items of wrong epenthesis are judged that words would not be pronounced in Japanese, which was significantly different from singleton and gemination ( $p < 0.001$ ) across all places of articulation. This suggests that L2 learners who know loanword phonology in Japanese know that appropriate epenthetic vowels should be inserted for the given phonological environment. This also suggests that learners are sensitive to distributional regularities in word-final consonants from English source words.

For NZE-speaking non-learners of Japanese, the effect of place of articulation is observed in an interesting way. While wrong epenthesis was found to be more disfavoured than singleton and gemination in the labial and velar context ( $p < 0.001$ ), the response pattern of wrong epenthesis for alveolar was significantly different from CVC but not from singleton and gemination. This suggests that non-learners prefer the [u] epenthetic vowel to [o] epenthetic vowel for the alveolar context. It could be attributed that non-learners develop epenthetic rules during their first exposure to the Japanese language, whereas detecting default epenthetic vowel patterns inhibits speakers from learning contextual epenthetic vowels. Thus, non-learners might find that pronunciation type with contextual epenthetic vowel [o] is marginal.

#### **4.5.1.2 Gradient phonological rules: Consonant gemination**

First, native Japanese speakers found gemination to be a more appropriate pronunciation than singleton in the voiceless context regardless of the place of articulation, which was significant ( $p < 0.001$ ). In the voiced context, singleton was significantly preferred for the labial and velar contexts ( $p < 0.001$ ). However, for the alveolar context gemination was found to be significantly better than singleton ( $p < 0.001$ ). It is not surprising that the final [d] is frequently geminated in nativised loanwords in Japanese (see §2.3.2.3). Thus, the effects of voicing type were found for singleton and gemination. The effects of POA for gemination in the voiced context are also found as alveolar [dd] > ( $p < 0.001$ ), velar [gg] > ( $p < 0.001$ ), and labial [bb]. These findings suggest that the listeners are sensitive to the stochastic distribution of word-final consonants in nativised loanwords.

Next, the results of NZE-speaking learners of Japanese are discussed in relation to phonological factors. L2 learners judge that gemination and singleton are words that would be pronounced in Japanese regardless of voicing types in the labial and velar contexts. There were no differences for ratings between singleton and gemination regardless of voicing type. The results demonstrate that while L2 learners apparently know Japanese syllable structures, they do not know the regularities of word-final gemination in relation to the voicing of source word

segments, at least in these contexts. On the other hand, in the alveolar context, the effects of pronunciation are found as gemination > ( $p < 0.05$  in the voiced context;  $p < 0.01$  in the voiceless context) singleton > ( $p < 0.001$ ) wrong epenthesis > ( $p < 0.001$ ) CVC. Thus, the response patterns in this context exhibit the observed nativised patterns in loanwords. Only in alveolar contexts shows that the observed pattern might be due to the frequency of alveolar geminates. The corpus data from the BCCWJ show that of the frequency of stop geminates in the first few hundred high frequency loanwords, [tt] and [dd] are higher in frequency than any other stop consonant for each voicing context. However, the effect of POA for gemination in the voiced context is not found; contrary to the prediction, the alveolar gemination received lower ratings than that of velar and labial ( $p < 0.001$ ).

As for NZE-speaking non-learners of Japanese, singleton and gemination were not significantly different across all places of articulation. The effect of voicing is not found at all. However, the effect of place of articulation is observed in an interesting way. While wrong epenthesis was found to be more disfavoured than singleton and gemination in the labial and velar context ( $p < 0.001$ ), the response pattern of wrong epenthesis for alveolar was significantly different from CVC, but not from that of singleton and gemination. This suggests that non-learners prefer the [u] epenthetic vowel to [o] epenthetic vowel for the alveolar context. Similar with L2 learners, it could be assumed that non-learners find epenthetic rules during their first exposure to the Japanese language, whereas detecting default epenthetic vowel patterns inhibit speakers from learning epenthetic vowel context. Thus, non-learners might find pronunciation type with contextual epenthetic vowel [o] is marginal.

#### **4.5.1.3 The effect of exposure to the target language (PC-JJ, PC-JNZ)**

As for extra-linguistic factors (PCA score for exposure to Japanese), both PC-JJ and PC-JNZ did not influence the response pattern as expected. That is, firstly, the effect of PC-JJ indicates that participants with higher exposure to Japanese in Japan tend to give a rating higher for stimuli in general, regardless of pronunciation types. Secondly, as for PC-JNZ, NZE-speaking learners of Japanese with less exposure to Japanese were more likely to rate singleton higher in the voiced context. This pattern would be expected for learners with higher exposure scores. However, as the score kept increasing, gemination tended towards higher ratings in the voiced context. In the voiceless context, while participants with less exposure to Japanese were not influenced by pronunciation types between singleton and gemination, participants with more exposure to Japanese were likely to rate gemination higher. This indicates that exposure to Japanese influences the perception of voiceless geminates. Since the L2 study context for

learners of Japanese is not immersive in Japan, benefits of exposure to Japanese is apparently limited. Exposed to a target language on daily basis both inside and outside of the class room not only mitigates access to L1 (Linck, Kroll, & Sunderman, 2009), but also increases opportunities to encounter and practice new words in the L2 without extra learning effort (Kojic-Sabo & Lightbown, 1999). This could be the reason that the effect of exposure was not detected.

On the other hand, NZE-speaking non-learners of Japanese with more exposure to Japanese in NZ were influenced by pronunciation types. That is, non-learners with more exposure to Japanese were more likely to rate gemination higher and CVC lower. These findings suggest that minimal experience with the target language affects statistical learning (Potter, Wang, & Saffran, 2017). Even though non-learners were not able to detect the target phonological regularities according to the phonological contexts from natural language environments, they seem to know Japanese phonology more than CV structure.

Considered together, findings from the well-formedness task answer to the research question stated, regarding **(RQ1)** whether a sublexicon phonology of a language is possible to learn from exposure to the target language, as it was predicted that the learning of phonological rules is possible without being taught. However, we could not observe that acquisition of phonological rules differs depending on the degree to which a learner is exposed to Japanese. At least within the current study, higher exposure to Japanese is not linked to having had a reasonable level of phonological knowledge.

Regarding **(RQ2)**, if any, what rules are implicitly learned? As predicted, L2 learners of Japanese are more likely to acquire epenthetic rules (i.e., categorical rules) than geminate rules (i.e., gradient rules) by exploiting their lexicon. However, language users are likely to overgeneralise an epenthetic rule in which [u] can be used in any context.

Regarding **(RQ 3)**, are L2 learners with high exposure to Japanese able to acquire fine-grained knowledge regarding the effects of voicing and place of articulation (POA)? While native Japanese speakers demonstrated gradient knowledge for voiced geminates, L2 learners were not influenced by voicing types or place of articulation.

Finally, regarding **(RQ4)**, do the language's overall statistical patterns influence learners' response patterns, and specifically are participants biased in responding with the most expected pattern in the language rather than with observed patterns in the Japanese loanword? In response to these questions, we found that singletons were selected frequently as geminates. However, the effect of sublexicon phonotactic was found for native Japanese speakers and L2 learners. Most importantly, the likelihood of ratings contrast between learners and non-learners



is quite different, indicating that learners acquire sublexicon phonology of Japanese to some extent.

#### **4.5.1.4 Rating for CVC**

Although the ratings for CVC are not the focus of the current study, let me briefly summarise the main findings. Considering native speakers of Japanese, while CVC received high ratings than wrong epenthesis ( $p < 0.001$ , in the labial, velar and voiceless alveolar contexts;  $p < 0.05$  in the voiced alveolar), the post-hoc analysis revealed that participants with more exposure to English in NZ were more likely to give lower ratings to CVC than participants with less exposure to English. That is, the high ratings are most likely due to perceptual confusion between singleton (CVCV) and CVC. These results are not contradicted, in that statistical learning in L1 affects our perception and production at several linguistic levels (Wilson & Davidson, 2013). As discussed in the pilot study section (see §3.3.2.1), listeners' perception of non-native sounds is constrained by their native language phonology (e.g., Berent, Steriade, Lennertz, & Vaknin, 2006; Dupoux et al., 2011), and consequently non-native listeners perceive an illusory vowel. These findings might indicate that the more people learn English the less likely they are to perceive illusory vowels after final consonants.

For learners of Japanese, when CVC items were presented, the participants confidently judged that words would not be pronounced in Japanese, which was significantly different from singleton and gemination ( $p < 0.001$ ) as well as wrong epenthesis. This result is expected as they learn the *hiragana/katakana* syllabary to write words in Japanese.

As for non-learners of Japanese, CVC received significantly lower ratings than other pronunciation across all place of articulations ( $p < 0.001$ ).

Next, the results of NZE-speaking learners of Japanese are discussed in relation to phonological factors.

#### **4.5.2 Discussion**

The most important findings in this study are that not only native speakers of Japanese but also L2 learners discern the structure of loanwords in Japanese from that of overall Japanese. They learn the sublexicon-specific structure from given complex input in the natural language setting.

Overall, native speakers of Japanese were able to detect stochastic distributional patterns in loanwords by attending to features of stop consonants in the coda position of the English source words. For native speakers of Japanese, all pronunciations are phonotactically

possible in their native language, with the exception of CVC and wrong epenthesis in alveolars. The question is why L2 learners could not detect gradient distributional patterns of geminates according to voicing types and place of articulation.

One possibility is that their vocabulary size for English loanwords is too small to detect these patterns by way of language experience. In order to detect specific phonological patterns, learners need to hear and use more words with these structures, and then they need to apply the knowledge to novel words (Edwards et al., 2004). The estimated loanword vocabulary size of learners of Japanese is around 300 words, which might be not enough to provide useful words that help learners to attend to types of regularities in words, and enable them to generalise structures to a new instance, according to phonological context.

Another possibility is the process of learning. Previous studies (e.g., Bulgarelli & Weiss, 2016; Pacton & Perruchet, 2008; Pacton, Sobaco, & Perruchet, 2015) reported that adult learners have difficulty learning various sets of regularities through the same inputs. For example, Pacton and Perruchet (2008) reported that when processing stimuli involving adjacent and nonadjacent digits, participants who were asked to focus on adjacent elements learned adjacent dependencies but did not learn nonadjacent dependencies. The other half of the participants who focused on nonadjacent elements showed they learned the other way around (i.e., nonadjacent dependency learning occurred but adjacent dependency learning did not occur). In a follow-up study, Pacton et al. (2015) made small changes to the previous task by adding the CVC syllables as another type of stimuli, along with sequences of digits and the new task did not require the selective processing of either adjacent or nonadjacent dependencies. As a result, irrespective of stimulus types (i.e., digits vs. syllables), participants learned adjacent dependencies significantly better than nonadjacent dependencies. Bulgarelli and Weiss (2016) argue that after learners have achieved robust learning for their first structure, they are less attentive to second structure (we discuss later this in §6.2.5 again). Such an account might explain the findings from the experiment.

First, the epenthetic rules were learnt in their own language environments. That is, using epenthetic vowels is necessary, for example, when learners need to write down their name from graphemic forms to *katakana* syllabary (e.g., クリス *kurisu* ‘Chris’) in Japanese. However, the quality of epenthetic vowels is dependent on the quality of the preceding consonant. Thus, learners need to pay attention to these regularities when they encounter English source words. The findings from the experiment suggest that learners were able to detect and track the epenthetic rules from their language experiences, however, response patterns indicate they

were not strongly confident in rejecting the illicit phonotactic sequences [tu] and [du]. This suggests that the default epenthetic vowel [u] was learnt first, and applied to the alveolar context as well as other contexts (i.e., labial and velar), even though the appropriate epenthetic vowels after alveolars is [o]. Since L2 learners rated wrong epenthesis with phonotactically licit sequences low in two other contexts, they apparently know the default epenthetic vowel [u]. This suggests that learners demonstrate their statistical learning towards the phonotactics for the loanword-specific patterns rather than overall phonotactics in Japanese. This interpretation was borne out by the findings that the learners accessed loanword phonotactics during the perceptual well-formedness experiment rather than accessing overall phonotactics of Japanese. Although learners have knowledge that [o] is the contextual epenthetic vowel after alveolars to some extent, their confidence was not as strong as judging default epenthetic vowel [u].

Similarly, for high ratings for gemination in voiced context, it is speculated that learners detect the regularities in the mono-syllabic English source words in which voiceless stops become geminates. The knowledge might block the learning of regularities depending on the voicing types of the stop consonants in the source word. Another possibility for the voicing types is that high frequent loanwords like /beddo/ ‘bed’ and /baggu/ ‘bag’ encourage learners to overgeneralise patterns to the labial context.

Other possible account is that although [tu] [du] are not common as another innovative variety such as [ti], [di], they are not actually illegal in the loanword stratum (see §2.3.2.1). That is, [tu] and [du] are attested sequences and recently acceptable for borrowings such as *tatuu* ‘tattoo’. Since the current study asked about pronunciation in Japanese when English words are borrowed in to Japanese, learners might have thought of such sequences as possible. However, in order to have such exemplars, L2 learners need to possess a certain vocabulary size.

The most interesting findings in relation to exposure levels to Japanese was that L2 learners with higher exposure to Japanese were less confident in judging CVC than learners with less exposure to Japanese. These results are compatible to the observed rating patterns of CVC by native Japanese speakers. From the current experiment design, it is difficult to say that an effect on the backward influence of L2 on L1 perception was detected. However, the findings might indicate that as more people exposure to a second language, it is more likely that they acquire expectations consistent with the structures of the second language. That is why in general, advanced L2 learners perform well in perceptual discrimination tasks. That is, while greater exposure to English presumably leads to enable native Japanese speakers to

distinguish between CVC and CVCV, speakers with lower exposure to the language cannot. Similarly, the auditory perception of native English speakers with greater exposure to Japanese might become increasingly similar to the perception of native Japanese speakers.

### 4.5.3 Remaining issues

This chapter has shown that it was possible for L2 learners to learn the sublexicon phonology of a language from exposure to the target language in natural language settings. Even though learners with high exposure to Japanese were not able to acquire fine-grained knowledge regarding the effects of voicing and place of articulation, they accessed loanword phonotactics rather than overall Japanese phonotactics during the experiment. However, their response patterns for singleton and gemination were not significantly different in the labial and velar contexts for both voicing types. L2 learners who do not find the regularities in their loanword lexicon might access statistical patterns in the overall lexicon for their responses, in other words their judgments were influenced by the most frequent phonotactic patterns in the language. In that case, singleton would be selected rather than geminates, as the overall frequency of geminates is much lower in the Japanese language (see §4.1.2.2).

However, the current response pattern in the findings leads to a fundamental question as to whether English-speaking learners of Japanese perceive differences between singletons and geminate consonants (e.g., /soku/ vs. /sokku/). In addition, native speakers of Japanese perceive differences between singletons and CVC (e.g., /soku/ vs. /sok/). In other words, the results of the current experiment might reflect perceptual confusion between non-native sound contrasts. A reasonable assumption is that the perceptual confusion arises from the fact that (1) While a consonant length contrast is phonemic and lexical in Japanese (e.g., *oto* ‘sound’ and *otto* ‘husband’), in English it is not. (2) While English allows syllable/word-final consonants such as ‘cat’ and ‘map’, Japanese does not. It is well known that speech perception is formed by the structure of the native language (e.g., Dupoux et al., 2011; Polka, 1992; Polka & Werker, 1994; Strange, Akahane-Yamada, Kubo, Trent, & Nishi, 2001; Werker & Tees, 1984). Therefore, non-native language listeners might have difficulty in perceiving the differences of the contrasts.

There is a consensus among researchers that individuals apply statistical learning to different levels of language (Romberg & Saffran, 2010). At the lowest level, statistical learning applies to categorisation of speech sounds (Romberg & Saffran, 2010). Phonetic information is categorised during the course of exposure to a language and using words, and “adults have

also abstract away a ‘phonological grammar’ of generalizations about where different phonetic categories are likely to occur” (Edwards et al., 2004, p. 422). Therefore, it is expected that because phonetic categorisation precedes other levels of language, without this process, statistical learning of phonological rules does not occur.

During the experiment, there were 240 stimuli that included four different pronunciation types and these were presented in random order to each participant. Since there was no direct comparison between two stimuli, it is possible that the learners of Japanese in the experiment might not have perceived differences between singleton and geminate consonant stimuli, whereas native speakers of Japanese also might have confused CVC and singleton stimuli. In order to explore this issue, a second experiment, which is the focus of the next chapter, was designed in a way to investigate the question of whether listeners tend to pay attention to differences between pairs of sounds that are very similar.

In addition, if specific stimuli are difficult to perceive, it might be due to differences in overall acoustic contrasts of CV sequences (e.g., differences between the acoustic properties of /ku/ and /to/). That is, the effect of acoustic salience in stimuli. This issue will be also explored as a plausible factor on perceptual confusion as a potential wider implication of the perceptual discrimination study.

## **4.6 Conclusion**

The current study attempted to answer the research questions by using well-formedness task across auditory stimuli of different pronunciations by comparing the response patterns of three groups: native Japanese speakers, NZE-speaking learners and non-learners of Japanese. The main findings are as follows:

1. Both native Japanese speakers and learners of Japanese access loanword phonotactics rather than overall phonotactics of Japanese for their responses. Thus, it is possible to learn a sublexicon phonology of a language through experience of using and/or passive exposure to the target language without being taught.

2. As predicted, epenthetic rules were learned by non-native Japanese speakers including non-learners of Japanese.

3. Members of the L2 learner group who had high exposure to Japanese were not able to acquire fine-grained knowledge regarding the effects of voicing and place of articulation.

L2 learners did not show different response patterns according to voicing types in the stimuli, while native Japanese speakers did reveal fine-grained knowledge regarding the effects of voicing and place of articulation.

4. Response patterns of non-learners show that they are sensitive to Japanese syllable structures, which varies dependent on their levels of exposure to Japanese.

Loanword phonology is a sublexicon phonology of Japanese phonology. These phonological regularities embedded in loanwords in Japanese are not usually taught in the language classroom. Therefore, the current study has analysed how L2 learners process and represent phonological regularities in relation to the coda status of English consonant, enabling us to better understand their ability of detecting distributional patterns in the linguistic properties.

# Chapter 5

## Auditory Memory Decision Task: Discriminating Non-native Segmental Contrasts in Spoken Word Lists

### 5.1 Introduction

This chapter conducts a perceptual experiment in order to provide insight into why singletons (i.e., CVCV) and geminates (i.e., CVCCV) in labial and velar contexts received equivalently high-ratings from NZE-speaking learners of Japanese in the first experiment, despite the fact that distributional pattern of geminates in the target structure (i.e., English CVC words) differ from that of singleton in loanwords. Providing an answer to this important question will aid interpretation of the results of the first experiment.

On the basis of previous studies and corpus data, we know that the stochastic patterns are such that voiceless geminates are more likely than voiced geminates in loanwords. Thus, geminates should have been rated higher than singletons in voiceless contexts (i.e., geminates > singletons). On the other hand, singletons should have been rated higher than geminates in voiced contexts (i.e., singletons > geminates). However, these patterns were not found in the learners' group. One possibility is that perceptual confusion between singletons and geminates was responsible for the results. That is, NZE-speaking learners of Japanese in the first experiment might have similarly perceived these contrasts. Therefore, the present work in this chapter investigates whether non-native speakers of a language are able to perceive consonant length contrasts that do not occur in their language.

As discussed in Chapter 4, response patterns in the findings lead to a fundamental question as to whether L1 English-L2 Japanese listeners can perceive differences between singletons and geminates /p/, /b/, /k/, /g/ followed by /u/. In addition, L1 Japanese listeners might have difficulty perceiving differences between singletons and CVC in the same phonological environments. For this group, CVC received surprisingly high ratings. These results could be attributed to the following two factors: (1) while consonant length contrast is phonemic and lexical in Japanese, in English it is not, and (2) while English allows syllable/word final consonants such as 'luck' and 'map', Japanese does not. Therefore, non-native language listeners might have difficulty perceiving the differences between the contrasts.

In the first experiment, the results in alveolar contexts by L1 English-L2 Japanese listeners suggest that the listeners probably perceive differences between singletons and geminates /t/ and /d/ followed by /o/ as the listeners showed a more gradient performance than between singletons and geminates /p/, /b/, /k/, /g/ followed by /u/. However, these listeners auditorily might not perceive contrastive differences between singletons and geminate /p/, /b/, /k/, /g/ followed by /u/.

Discrimination accuracy on these types of contrasts varies in previous studies, depending on the tasks and their conditions (Gerrits & Schouten, 2004). While many studies argue that acquisition of non-native sound length contrasts are difficult (e.g., Altmann, Berger, & Braun, 2012; Hirata, 2004; Hirata, Whitehurst, & Cullings, 2007; Tajima, Kato, Rothwell, Akahane-Yamada, & Munhall, 2008), some studies report that even naïve participants without exposure to a target language show high accuracy when discriminating non-native length contrasts (e.g., Asano, 2018 (for short inter-stimulus-interval condition); Hayes, 2001; Hisagi & Strange, 2011). Studies with high accuracy results often resulted from same-different (AX) discrimination tasks during which subjects listen to pairs of the test stimuli and determine whether the two stimuli were the same or different. The task is straightforward and relatively easy as participants are able to compare stimuli directly and to decide that two stimuli are somewhat different. The AX experiment is designed to encourage listeners to access auditory information available (Gerrits & Schouten, 2004; Pisoni, 1973). Although discrimination is possible when minimal pairs occur adjacently, contrastive words rarely appear next to each other in real communicative situations. In fact, in AXB discrimination experiments, naïve English listeners' performance for three types of Japanese length contrasts was significantly poorer when the contrast types were presented randomly, than when they were presented in separate blocks (Hisagi & Strange, 2011).

For L2 learners, listening environments of everyday oral communication are not always optimized, and a listener's perception of non-native sound contrasts is vulnerable. Other aspects add to this vulnerability, as indicated during the first experiment in Chapter 4, where 240 stimuli including four different pronunciation types for each word were presented in random order to each participant. Since there was no direct comparison between two stimuli, it is possible that the learners of Japanese in the experiment might not have perceived differences between singleton and geminate consonant stimuli, whereas native Japanese speakers also might have confused CVC and singleton stimuli. Perception of non-native sounds is known to be susceptible when task demands are high (e.g., Asano, 2018). It would be expected that increasing working memory load degrades the performance of spoken word



discrimination (Werker & Logan, 1985). Thus, the primary aim of this chapter is to address the question of whether non-native speakers of a language have phonological awareness of non-native sound contrast with high memory demand. I will try to determine why singletons /p/, /b/, /k/, /g/ followed by /u/ received equivalently high-ratings as the counterpart of geminates from NZE-speaking learners of Japanese in the first experiment, as well as why CVC in which the last consonants were /p/, /b/, /k/, /g/ received relatively high-ratings as the counterparts of CVCV stimuli from L1 Japanese listeners. This chapter will address whether those results arise from perceptual confusion.

Thus, this dissertation also concerns the effects of acoustic similarity of segments on perceptual discrimination. If there is perceptual confusion, it might be due to differences in phonetic salience of particular acoustic signals. Therefore, this study will address the potential source of perceptual confusion across contexts.

Thus, this chapter examines the degree to which non-native speakers can perceive contrasts that do not occur in their native language – singleton/geminate contrasts for English speakers, and CV/CVC contrasts for Japanese speakers. It also investigates the degree to which success in this task is mediated by phonetic salience of the particular contrast, and by the individual's previous language experience.

## **5.2 Background for Memory Decision Task**

In this section, previous studies relevant to the research questions for the first objective will be reviewed. Prior studies regarding perception of non-native length contrasts argue that attention control affects listeners' perception of non-native sound contrast (e.g., Asano, 2018; Hisagi & Strange, 2011; Porretta & Tucker, 2015), and the effect of task demands plays an important role for the perception of non-native contrastive sounds. This section begins to explore these influential factors on speech perception.

### **5.2.1 The effects of attention control for discrimination of non-native length contrasts**

In terms of the effect of participants' attention, perception of non-native sounds is known to be more enhanced by giving participants an optimal listening condition in which they can pay attention to the contrast (Asano, 2018; Guion & Pederson, 2007; Hisagi & Strange, 2011; Pederson & Guion-Anderson, 2010; Porretta & Tucker, 2015). Attention control is the cognitive ability to shift efficient attention between task-relevant and task-irrelevant information (Isaacs & Trofimovich, 2011; Rosen & Engle, 1998).

For instance, Hisagi and Strange (2011) found the effects of attentional variables on naïve American listeners' perception of Japanese contrasts of vowel length (e.g., kiro vs. kiiro), consonant length (e.g., kite vs. kitte) and syllable number/length (e.g., k<sup>j</sup>oo vs. kijoo) in AXB discrimination tasks. Although there were no significant overall differences in relative difficulty across contrast types, listeners' discrimination performance was significantly different depending on how stimuli were presented. In comparison to when the three contrast types were presented in separate blocks with detailed instructions about the nature of the contrasts, listeners' performance was significantly poorer when the identical stimuli were presented randomly, intermixed with no specific instructions. Hisagi and Strange attributed the results to the factor that listeners' attention had not been directed to temporal cues differentiating the contrasts.

Similarly, Porretta and Tucker (2015) conducted experiments in which there were two groups of native speakers of English and one group of native speakers of Finnish. For native speakers of English, one group (N=20) was informed that they would hear foreign words whereas the other group (N=20) was informed they would hear Finnish nonwords that have a consonant length contrast. As a result, participants' ability to perceive a non-native consonantal length contrast in the second group increased significantly over that of the no-instruction group in an AX discrimination task as well as a forced-choice identification task. These studies show that the knowledge of consonant length contrast assists participants in paying attention to consonant duration as an acoustic temporal cue distinguishing singletons from geminates across experiments. As it turned out, perceptual ability of non-native contrast was enhanced. Explicit awareness of specific information in input plays a role in perceptual discrimination.

The effect of attention is also found in another study. Asano (2018) tested the effects of memory load by increasing the inter-stimulus-interval (ISI, 300ms vs. 2500ms) and the effects of attention control by manipulating pitch conditions of stimuli (high-flat pitch vs. high-falling pitch) when discriminating between non-native consonant length contrasts. Participants were German-speaking learners of Japanese (N=48), German non-learners (N=24) and native speakers of Japanese (N=24). Participants were instructed to pay attention to duration and asked to judge whether the two stimuli were the same or different in terms of their duration. The results showed high discrimination abilities with the shortest ISI and flat pitch conditions across all groups, and even the non-learners were able to discriminate between the contrasts. When the ISI was 2500ms, the performance of non-learners was weakened, while two other groups were not influenced by the increased memory load. This suggests that when the non-learners were required to tap into phonological processing during perception, they faced

difficulties as they relied on phonetic levels of information in the stimuli. On the other hand, the sensitivity of both learners and non-learners was strongly influenced by the task-irrelevant falling pitch condition, in contrast to the flat pitch condition. Asano (2018) argues that “difficulties arose for both non-native listener groups in ignoring task-irrelevant pitch, suggesting the complications in automatizing L2 processing even after being exposed to or establishing the L2 categories” (p. 426).

Similarly, a perception study by Hardison and Saigo (2010) found that presenting stimuli in a carrier phrase hinders identification accuracy of geminates in contrast to identification in isolation by English-speaking learners of Japanese (N=85), regardless of their Japanese proficiency. These findings indicate that difficulty in perceiving non-native contrasts still remains, unless it occurs under optimal listening conditions so that L2 learners can pay attention to the target stimuli.

In summary, the studies described above have shown that attention control plays an important role in discriminating non-native length contrasts. In particular, when given explicit instructions about segmental length information, participants can generally direct their attention to the temporal durational cue that differentiates between singleton and geminate consonants. Thus, their discrimination accuracy improves. However, when task demand is high or without explicit instructions regarding input, listeners cannot attune their attention to particular auditory and phonetic information. Thus, listeners are less likely to access cues differentiating the contrasts in making perceptual judgments. These studies suggest that NZE-speaking learners of Japanese in the first experiment might not have paid attention to segments differentiating the contrasts. Therefore, listeners might have not been able to successfully perceive differences between singletons and geminates for learners, and between singletons and CVC for Japanese. With this in mind, I will next review literature about memory load and information processing which explain why listeners have difficulty in discriminating between stimulus contrasts in the case of high memory loads. Then, literature particularly pertinent to the CVC case is reviewed in relation to the effects of acoustic similarity of segments on perceptual discrimination.

### **5.2.2 Task demand, memory load, and speech processing**

Another important factor related to speech perception is that different experimental tasks demand different types of speech processing. According to the framework of processing factors in speech perception, listeners use different speech processing strategies according to experimental conditions (Cowan & Morse, 1986; Crowder, 1982; Gerrits & Schouten, 2004;

Johnson, 2004; Pisoni, 1973; Schouten & van Hessen, 1992; van Hessen & Schouten, 1992; Werker & Logan, 1985; Werker & Tees, 1984). Listeners need to rely on language-specific phonemic/phonological information during high memory demand, whereas during low memory demands such as short ISI, listeners rely on auditory or phonetic information depending upon task conditions (Werker & Logan, 1985). This is attributed to the rapid decay of auditory information. Auditory memory traces become less reliable with increasing working memory load and, as a result, an auditory comparison of stimulus information is not possible and listeners need to rely on categories they assign to the stimuli.

In the first experiment, a well-formedness judgement task was employed to find out listeners' phonological knowledge. It was not designed to give access to auditory or phonetic information during the task. Listeners were expected to access their long-term memory rather than short-term memory in making judgments. That is, listeners required phonemic processing, through which they must rely on their phonological knowledge about loanwords which involve consonant length contrasts. If learners did so, the well-formedness ratings would then display the observed phonotactic patterns in nativised loanwords, but they did not. In actual fact, singletons and geminates received similar ratings by L1 English and L2 Japanese listeners except for alveolar contexts. Thus, during the first experiment, listeners might have not been able to tap into their phonological knowledge or might not have had any relevant phonological knowledge to start with. Therefore, singleton and geminate consonants were not perceived categorically due to perceptual confusion between those sounds. However, another case is also probable, in that even if the listeners discriminated the consonant contrasts, it is possible that they rated both stimuli highly. This would occur if learners had clear knowledge of phonological categories, but had not generalised the target loanword pattern from inputs they encountered. To evaluate these possibilities, the present experiment investigates whether they are able to perceive contrasting singletons and geminates. This is the main purpose of this chapter which will help to interpret the first experiment.

As seen above, auditory information is available during low memory demand tasks, such as an AX discrimination task. In such cases, even naïve listeners are able to discriminate non-native sound contrasts. Considering these facts, the present experiment explores whether non-native speakers of a language are able to perceive consonant length contrasts that do not occur in their language.

### 5.2.3 The role of phonetic salience on non-native perception of contrast

In addition to attention control, acoustic information of segments is important for perceptual discrimination. In this section, previous studies relevant to the question of phonetic salience will be reviewed. The perception of phonological contrasts varies from one context to another because certain sequences are phonetically clearer than others (e.g., Polka, 1991, 1992). In relation to the first experiment, if specific stimuli are difficult to perceive, it might be due to differences in overall acoustic contrasts of CV sequences (e.g., differences between the acoustic properties of /ku/ and /to/). This is a potential wider implication of the study, which is more interested in discrimination accuracy in light of the wider literature, as opposed to helping interpret the first experiment. Prior studies regarding the perception of non-native structure suggest that relative salience of the acoustic dimension in stimuli might play an important role for listeners' perception of non-native sound contrasts. I considered findings from previous studies on the perception of non-native sound contrasts by both native speakers of Japanese (e.g., Dupoux et al., 1999; Dupoux et al., 2011; Monahan et al., 2009) and native speakers of English (e.g., Hardison & Saigo, 2010).

Research on the perception of consonant clusters by Japanese listeners showed that they have a tendency to perceive an unrepresented [u] between two consonants regardless of voicing types of the preceding consonants (Dupoux et al., 1999; Dupoux et al., 2011; Monahan et al., 2009). Dupoux et al. (1999, 2011) conclude that the perception of non-native structures was constrained by Japanese phonotactics. That is, the 'mishearing' of [u] suggests perceptual restoration of a phoneme based on what listeners expect to be there. Phonemic restoration is a phenomenon in which listeners hear a missing speech segment illusory by perceptual filling-in (Shahin, Bishop, & Miller, 2009; Warren, 1970). Importantly, perceptual restoration is interact with listeners' L1 background (Ishida & Arai, 2015; Kashino, Wieringen, & Pols, 1992).

However, vowel epenthesis in Japanese is closely linked to the phonological context which complies with Japanese phonotactics. The contextually appropriate epenthetic vowel after alveolar stops is [o] (see §2.3.2). Interestingly, Japanese listeners were able to discriminate alveolar VCCV sequences from similar sequences with either a medial [u] or [o] (e.g., /etma/ vs. /etoma/, /etuma/), whereas the same listeners perceive an illusory [u] in velar VCCV contexts (e.g., /ekma/ vs. /ekuma/) (Monahan et al., 2009). In fact, in a series of perceptual experiments by Dupoux et al. (1999, 2001, 2011), all stimulus clusters in words given to Japanese listeners attract an epenthetic vowel [u] by their phonological context, if the listeners perceive an illusory epenthesis. If Japanese phonotactics solely led listeners'

perception, all closed syllables would be perceived as open syllables by inserting an epenthetic vowel, regardless of preceding consonants. These findings have raised implications for the potential role of acoustic information in explaining perceptual epenthesis, because [o] is more phonetically salient than [u]. Since [u] is considered to be the shortest vowel and the most susceptible to weakening and deletion among the five Japanese vowels (Hirayama, 2003; Kubozono, 2015; Sagisaka & Tokuhara, 1984 as cited in Irwin, 2011; Shoji & Shoji, 2014), common properties of perceptually weak vowels, the presence/absence of [u] might confuse differentiation between two consonants. On the other hand, [o] has higher sonority than that of [u], and so if [o] is not presented, listeners might readily notice. Similarly, an appropriate epenthetic vowel [u] was found after English coda [m] in other studies (Aoyama, 2003; Kilpatrick, Kawahara, Bundgaard-Nielsen, Baker, & Fletcher, 2018). Interestingly, English coda [m] also elicited [n] which is allowed in Japanese syllabic patterns (Kilpatrick et al., 2018). Thus, these studies suggest that while Japanese syllable structure constrains the perception by Japanese listeners, phonetic salience seems to play a role.

Thus, if Japanese listeners' phonological/phonotactic knowledge leads them to believe that a final consonant is impossible, they would show difficulty in discriminating contrasts of /ku/ vs. /k/ rather than /to/ vs. /t/. However, in the above studies, the perceptual confusability comes in word-medial consonant clusters. It is currently unknown whether Japanese listeners perceive an epenthetic vowel in word-final position for the velar context, as may have happened in the first experiment.

The effect of phonetic information is also assessed in English listeners' perception on consonant length contrasts in Japanese. Study of a forced-choice identification task by Hardison and Saigo (2010) showed that greater consonant-vowel sonority difference facilitated perception for geminates by L1 English-L2 Japanese listeners at three proficiency levels: beginner, low-intermediate, and advanced level. Stimuli consisted of the medial consonants /s/, /t/, and /k/, followed by /a/ and /u/. When stimuli (e.g., /sasu/, /sassu/, /sasa/, /sassa/, /saku/, /sakku/, etc.) were presented in carrier sentences, L1 English-L2 Japanese listeners showed most difficulty in identifying [ssu] as geminates, followed by [ssa], and additionally they showed high accuracy on [tta], and [kka] followed by [ttu][kku] (the stops /t/ and /k/ had similar accuracy patterns across all groups). Hardison and Saigo argue that greater sonority differences between geminate consonants and post-consonant vowels create better perceptual distance as acoustic cues, facilitating accurate identification. In the isolation condition, [sassa] was generally perceived with higher accuracy than [sattu/sakku]. Thus, the difference acoustic salience between the pair of segments might play a part in perception on length contrasts as

well as the acoustic salience of a segment.

Thus, in the first experiment, it is possible that NZE-speaking learners of Japanese face difficulty in perceiving contrasts between singleton and geminate /k/ or /g/ followed by /u/ than between singleton and geminate /t/ or /d/ followed by /o/. Therefore, the present experiment is designed to evaluate whether the results in the first experiment are due to perceptual confusion between non-native sound contrasts in the particular phonological environment. Especially so, as the current experiment aimed to investigate whether listeners can auditorily perceive contrasts between singleton and geminate /k/, /g/ followed by /u/, and between singleton and geminate /t/, /d/ followed by /o/ based on the first experiment. If the findings of the first experiment are influenced by perception effects, then, L1 English-L2 Japanese listeners should have perceptual confusion of non-native sound contrasts with the vowel /u/.

#### **5.2.4 Current study**

In order to explore the objectives mentioned above, an auditory memory decision task was chosen as a way to examine whether non-native speakers of Japanese are sensitive to a non-native consonant length contrast (CVCV-CVCCV), and non-native speakers of English are sensitive to sound pairs of CVC and CVCV. Similar to the first experiment in Chapter 4, stimulus words include singleton and geminate /t/, /d/ followed by /o/ and singleton and geminate /k/, /g/ followed by /u/. Understanding which sequences (i.e., [do], [to], [gu], [ku]) are more confusable with their CVC sequence counterparts (i.e., [CVdo] vs. [CVd], [CVgu] vs. [CVg]) enables us to identify any effect of phonetic salience of the segment. All phonemes are native to both English and Japanese. Because a similar response pattern was observed in the labial and velar contexts across all groups in the first experiment, labial stop consonants were not included as stimuli in the current study.

The experiment was presented as being about word memory but listeners were not told to focus on the precise pronunciation of words. This is because, as discussed in §5.2.1, when simple information about sound patterns is given as instructions, participants' ability to perceive non-native consonant length contrast is more likely to significantly increase than when they lacked such information (e.g., Porretta & Tucker, 2015). Thus, attention to 'words' rather than particular sounds was needed to maintain an experimental condition consistent with the first experiment.

Participants were asked to perform an auditory memory decision task on lists of stimuli that contained pairs of non-native sound contrasts. Participants were informed that they would

hear different words but not given the information that some words have closely related sound patterns. They heard a list of four words followed by a beep. After the beep, another word would be heard. For example, a participant heard /hapa/, /detto/, /ruku/, and /gate/ followed by a 300ms beep. Then the counterpart of the sound contrast pair /deto/ was heard. They were asked to decide whether the word after the beep was in the list of words before the beep by clicking ‘Yes’ or ‘No’ buttons on the computer screen. Some trials contained sound contrast pairs for which the correct answer (assuming the contrast is heard) would be ‘No’ whereas other trials contained non-contrast pairs for which the correct answer would be ‘Yes’. The ‘same’ tokens were different recordings of the same word.

Thus, the experimental design is intended to motivate participants to focus their attention on memorising a short list of words. This enables us to examine the actual learners’ perception in the first experiment, more analogous to everyday life or running speech than an AX discrimination task. Presumably, auditory memory would decay by the time the last word is presented because participants need to hold the word list in memory until they hear the last word to compare. Hence, the task prevents participants from relying solely on auditory information, forcing them to rely on phonemic/phonological/word-based information in speech perception.

The performance was compared across three groups: L1 Japanese listeners, L1 English-L2 Japanese listeners (i.e., learners of Japanese), and monolingual English listeners. Assigning the same experimental condition to all participants allowed for the comparison of response patterns across different groups. As in the first experiment, levels of exposure to Japanese/English were measured using PCA. The current study therefore seeks to contribute to a better understanding of the effect of phonological knowledge on recognition memory and perceptibility for spoken non-native phonemic contrast in relation to language exposure. The study sheds light on phonological awareness of native/non-native sound contrasts by comparing three different groups using a novel method. The details of the methodology will be discussed further in §5.3.

### **5.2.5 Predictions**

As clarified above, this chapter aims to examine the degree to which non-native listeners can perceive contrasts that do not occur in their native language – singleton/geminate contrasts for English listeners, and CV/CVC contrasts for Japanese listeners. It also investigates the degree to which success in this task is mediated by phonetic salience of the particular contrast, and by



the individual's previous language experience. Specifically, the present study investigates whether under the demand of memory load, listeners can auditorily perceive contrasts between singleton and geminate /k/, /g/ followed by /u/, and between singleton and geminate /t/, /d/ followed by /o/ in order to explain the findings of the first experiment. The processing strategies by Werker and Logan (1985) would predict that the memory decision task should lead to lower accuracy than show in previous studies because the task condition is more similar to everyday oral communication.

### **Prediction 1a: Discrimination of singleton – geminate contrasts**

(1) As for singleton-geminate contrasts, geminate stimuli are phonemic/phonological categories for native Japanese listeners, so that they would demonstrate phonemic perception of discrimination between the two stimuli.

(2) On the other hand, considering L1 English-L2 Japanese listeners, some listeners would show a sensitivity to phonemic distinctions between geminate stimuli, although the stimuli are not-native language phonological categories. If they could not tap into the phonemic/phonological information, they would show difficulty in discriminating between the stimulus contrasts.

(3) Monolingual English listeners would not be able to discriminate between consonants varying in duration as the listeners do not have phonemic categories according to consonant length.

Therefore, it is predicted that performance across language groups would differ.

### **Prediction 1b: Mediating effects of phonetic salience on perception of geminates**

If specific stimuli were difficult to perceive, this might be due to differences in overall acoustic contrasts of CV sequences (e.g., differences between the acoustic properties of /ku/ and /to/). That is, perception of phonological contrasts varies from one to another because certain sequences are phonetically clearer (e.g., Polka, 1991, 1992).

(1) Regardless of groups, if listeners show difficulty perceiving contrastive pairs, singleton and geminate /k/, /g/ followed by /u/ (i.e., velar stimuli) would be discriminated less well compared to singleton and geminate /t/, /d/ followed by /o/, because the velar context is less phonetically salient than the alveolar contexts. In this case, an effect of POA would be detected.

(2) If listeners are quite sensitive to acoustic signals, contrasts in voiced stimuli are more likely to be discriminated better than contrasts in voiceless stimuli. This is based on the degree of

acoustic difference in the voiced and voiceless stimuli; mean ratios of geminate to single closure (GC/SC) are about twice for voiced stimuli (acoustic details of stimuli will be discussed in §5.3.3). In this case, an effect of voicing type would be detected.

**Prediction 1c: The effects of the degree of exposure of the individuals to the target language on geminate perception**

Greater exposure to a target language presumably leads to an increased ability to identify the relevant phonological contrasts. Thus, levels of exposure to Japanese/English might affect listeners' performance. Hence, it would be predicted that the performance of L1 English-L2 Japanese listeners with high exposure to Japanese might be better than that of L1 English-L2 Japanese listeners with low exposure to Japanese.

**Prediction 2a: Discrimination of singleton – CVC contrasts**

(1) As for singleton-CVC contrasts, monolingual English listeners will demonstrate phonemic perception of CVC stimuli that are within their L1 phonological categories.

(2) As for Japanese listeners, if their phonological/phonotactic knowledge leads them to believe that a final consonant is impossible, they would show difficulty in discriminating contrasts in the stimuli.

Therefore, it is predicted that performance across language groups would differ.

**Prediction 2b: Mediating effects of phonetic salience on perception of CVC**

Regardless of groups, if listeners show difficulty perceiving contrastive pairs, singleton and CVC in the velar context would be discriminated less well compared to in the alveolar context, because the velar context is less phonetically salient than the alveolar contexts. In this case, an effect of POA would be detected.

**Prediction 2c: The effects of the degree of exposure of the individuals to the target language on CVC perception**

(1) It would be predicted that the performance of L1 English-L2 Japanese listeners with high exposure to Japanese might be lower than that of L1 English-L2 Japanese listeners with low exposure to Japanese. This is based on the assumption that as more people are exposed to a second language, it is more likely that they acquire expectations consistent with the structures of the second language. That is why in general, advanced L2 learners perform well in perceptual discrimination tasks.

(2) Similarly, it would be also predicted that the performance of native Japanese listeners with high exposure to English might be better at perceiving between singleton and CVC than that of native Japanese listeners with low exposure to English.

If all predictions are true, this will supply conclusive information regarding the first experiment as to whether L2 learners discriminated the consonant contrasts (i.e., singleton/geminate) but they rated both stimuli highly. This would occur even if learners had clear knowledge of phonological categories, but they had not generalised the target loanword pattern from inputs they encountered.

## **5.3 Research Design and Methodology**<sup>8</sup>

### **5.3.1 Materials**

Stimulus materials consisted of four sets of 10 minimal triplets, giving a total of 120 target words. Additionally, another 120 words as fillers were created, consisting of disyllabic CVCV non-Japanese words (see Appendix F). All the first vowels in the target words were lax vowels, and second consonants were selected from either alveolar or velar stop consonants (i.e., /t/, /d/, /k/, /g/) in order to match the previous experiment. When the second consonants in the words were the velar stops /k/, /g/, the following vowel was /u/ (e.g., /keku/, /geku/, /segu/, /migu/). When the second consonants in the words were the alveolar stops /t/, /d/, the following vowel was always /o/ (e.g., /kuto/, /keto/, /suto/, /nedo/). Those following vowels are the legal epenthetic vowels after alveolar stops. That is, there were no stimulus words with /k/, /g/ followed by /o/ or with /t/, /d/ followed by /u/. Thus, four sets of word lists that end with two different voicing types and two places of articulation were created. The use of this methodology enables us to look more closely at the effect in the previous task and the types of consonants on recognition memory for spoken words.

All stimuli were nonwords or existing words with low frequency, because studies show that lexical knowledge influences phonemic identification (Frauenfelder, Segui, & Dijkstra, 1990; Ganong, 1980). In addition, using nonwords or low frequency words reduces the disadvantage non-native speakers' experience when having no knowledge of lexical information in words. For the low frequency words, similar to the first experiment, the BCCWJ (National Institute for Japanese Language and Linguistics, 2011) was used to ensure that the

---

<sup>8</sup> This experiment was pre-registered using AsPredicted (<https://aspredicted.org/index.php>). The ID is #27006. The preregistration document appears in Appendix E.

wpm of words be lower than 1.92 (i.e., frequency is 1.92 words per million). Stimulus words consisted of minimal triplets such as /keto/ (CVCV), /ketto/ (CVCCV) and /ket/ (CVC). CVCV structured non-Japanese words were created first, and geminate and monosyllabic words were then created. Therefore, some CVC words exist as English real words, as shown in Table 5.1. From this point, contrast pairs containing alveolar stops + /o/ will be called *alveolar stimuli*, and contrast pairs containing velar stops + /u/ will be called *velar stimuli*.

Table 5.1 Example of stimuli

	Reference nonword	Singleton	Geminate	CVC
Alveolar	keto	/keto/	/ketto/	/ket/
	sudo	/sudo/	/suddo/	/sud/
Velar	keku	/keku/	/kekku/	/kek/
	segu	/segu/	/seggu/	/seg/

### 5.3.2 Audio stimuli

The stimuli for a perception experiment were created by recording a native speaker of Japanese who was the same speaker as the first experiment. Only one speaker was selected to avoid any effect of talker variability (e.g., Bradlow, Nygaard, & Pisoni, 1999; Martin, Mullennix, Pisoni, & Summers, 1989; Mullennix, Pisoni, & Martin, 1989). The recording took place in a sound-attenuated room at the University of Canterbury, using Audacity and USBpre2 with 44,100 samples/s, 16 bit/s and a Beyerdynamic head-mounted microphone. The stimulus speaker's participation was approved by the University of Canterbury Human Ethics Committee (under application number HEC 2019/27/LR-PS).

Each word was presented in Japanese and produced using PowerPoint slides. The speaker read a set of slides which contained one word per slide. Each word was presented three times, allowing the use of different tokens of the same pronunciation type. Each slide was presented 2.5 seconds apart to help maintain the same tempo for the reading of each word. A tone pattern of all target words was a HL/HLL sequence. This is the default accent for loanwords in Japanese (Kindaichi & Akinaga, 2014), and is commonly used for singleton-geminate discrimination experiments (e.g., Asano, 2018; Hardison & Saigo, 2010). Then, production recordings were analysed acoustically using Praat phonetic software (Boersma & Weenink, 2018). Giving consideration to clarity of production and the duration of words, the marginal recording was manually removed from the three options. It should be noted that the

CVC stimuli contained stop release bursts. Some audio files for CVC and CVCV were heard by colleagues in order to validate whether intended sounds were produced or not.

### 5.3.3 Acoustic characteristics of the audio stimuli

As before, in order to determine the acoustic characteristics of the stimuli, all vowels in the audio stimuli were analysed using Praat (Boersma & Weenink, 2018). The duration of each target vowel was measured and the mean values for F1, F2, and F3 for the stimulus vowels were extracted using a custom Praat script. All measurements were taken at the midpoint of the marked segment. All extracted formants were checked manually to ensure the validity of the values. Note that /i/ and /u/ are not devoiced in the environments of preceding voiceless consonants or between two voiceless consonants.

First, vowel formants are considered. The mean values for F1, F2 and F3 for the stimulus vowels and their standard deviation (SD) are shown in Table 5.2. As these are all from a single speaker, formant values were not normalised.

Table 5.2 Mean F1/ F2/F3 formant values and standard deviations for voiced and voiceless contexts

Number of Token	Vowel	Symbol for Plot	Position in Word	F1		F2		F3	
				mean	SD	mean	SD	mean	SD
16	[a]	a	medial	842	72	1670	67	2992	105
93	[e]	e	medial	519	56	2319	82	3031	100
22	[i]	i	medial	275	14	2769	36	3584	278
5	[o]	o	medial	550	74	1069	132	3334	74
75	[o]	O	final	451	30	939	58	3008	236
78	[u]	u	medial	343	39	1481	139	3062	150
72	[u]	U	final	379	31	1550	54	2973	56

Figure 5.1 shows the individual vowels from the speaker. For this context, the vowel /o/ in word-final position is indicated as ‘O’ and the vowel /u/ in word final position as ‘U’. As can be seen, vowels in the plots do not overlap with each of the other five vowel categories. The F1/F2 space for the stimulus vowels is consistent with the first experiment.



Figure 5.1 F1 and F2 values for individual vowels from the stimulus speaker.

Secondly, vowel duration is considered. Table 5.3 shows descriptive statistics for vowels produced by the speaker. The standard deviation of vowel duration ranges from 12 to 23 ms, consistent across vowels. The ranking of vowels from longest to shortest is [o] [u] in the word-final position followed by [o], [a], [e], [i] [u] in word-medial position. The results are fairly consistent with earlier vowel duration studies except [o] which is generally shorter than [a] and [e] (Campbell, 1992; Han, 1962 cited in Shoji & Shoji, 2014; Yoshida, 2006). This is possibly due to numbers of medial [o] in the current study being very small. As before, although [u] is acknowledged as the shortest vowel in Japanese, cross-linguistically vowels in utterance final position are normally longer than vowels in non-final positions (Johnson & Martin, 2001), and final mora is lengthened in Japanese (Kaiki et al., 1990).

Table 5.3 The number of token counts, the mean and SD in ms. of the five vowels

Number of Token	Vowel	Symbol for Plot	Position in Word	Vowel Duration (ms)	
				mean	SD
16	[a]	a	medial	122	20
93	[e]	e	medial	108	23
22	[i]	i	medial	104	21
5	[o]	o	medial	133	12
75	[o]	O	final	165	17
78	[u]	u	medial	93	21
72	[u]	U	final	158	20

According to previous studies (e.g., Han, 1994; Hirata, 2007; Hirata & Whiton, 2005; Kawahara, 2015 for a summary of Japanese geminates), acoustic properties are very important factors for distinction of single and geminate voiceless stops in Japanese. In general, the closure duration of a voiceless geminate is more than twice as long as that of a voiceless singleton (Kawahara, 2015). Following previous studies, VOT was not included for the closure duration measurement because it does not affect the perception of the geminate-singleton contrast (Han, 1992). In the current study, the ratio of geminate to single closure duration is followed: [t]: [tt] 1:2.45, [k]: [kk] 1:2.42, [d]: [dd] 1:5.12, [g]: [gg] 1:4.22. The ratio of voiced consonants was greater than that of voiceless, which is consistent with the findings in the previous studies on closure durations of singleton and geminate stops (e.g., Homma, 1981). In addition, some acoustic properties are invariant for distinguishing singletons from geminates regardless of speech rate and speakers (Hirata, 2007; Hirata & Whiton, 2005; Idemaru & Guion-Anderson, 2010). These studies tested to see if accurate acoustic measures can identify singleton and geminate categories. In general, relational measures predict higher accuracy percentages for identifying the intended category than raw durational values do. Durational ratios of closure to word (C/W) (Hirata, 2007; Hirata & Whiton, 2005; Idemaru & Guion-Anderson, 2010) and closure to the following vowels (C/V2) (Hirata & Whiton, 2005; Idemaru & Guion-Anderson, 2010) provide accurate identification of consonant length between singleton and geminate. For example, when the optimal boundary value of C/W was 0.35, the categories were accurately identified at least 95% of the time (Hirata & Whiton, 2005). These invariant parameters of the stimulus speaker were measured for the current study, and data summarised in Table 5.4 and plotted in Figure 5.2-4 for voiceless in the left panels and for voiced stimuli in the right panels across place of articulation.

In Table 5.4, values in parentheses indicate standard deviations. In the figures, the black dots indicate the mean value. Those durational parameters do not deviate from the data of Hirata and Whiton (2005), as they claim the acoustic parameters are stable across different speaking rates and speakers. Note that the current data for voiced consonants are not directly comparable with the previous studies due to no available data for voiced consonants. Since the data clearly make distinction between singletons and geminates, these values were not statistically analysed.

Table 5.4 Mean duration (ms) of word, stop closure and mean ratios of geminate to single closure (GC/SC), closure to word (C/W), and closure to following vowel (C/V2)

	[t]		[k]		[d]		[g]	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Word Duration								
Singleton	436	(18)	466	(34)	433	(54)	467	(37)
Geminate	672	(35)	678	(38)	675	(63)	702	(46)
Closure Duration								
Singleton	139	(12)	139	(14)	62	(11)	76	(12)
Geminate	339	(33)	335	(26)	309	(21)	315	(23)
GC/SC Ratio	2.45	(0.27)	2.42	(0.23)	5.12	(1.19)	4.22	(0.61)
C/W Ratio								
Singleton	0.32	(0.03)	0.30	(0.03)	0.15	(0.03)	0.16	(0.03)
Geminate	0.51	(0.05)	0.49	(0.04)	0.46	(0.04)	0.45	(0.04)
C/V2 Ratio								
Singleton	0.90	(0.11)	0.99	(0.12)	0.36	(0.06)	0.43	(0.08)
Geminate	2.19	(0.22)	2.39	(0.23)	1.78	(0.21)	1.96	(0.31)

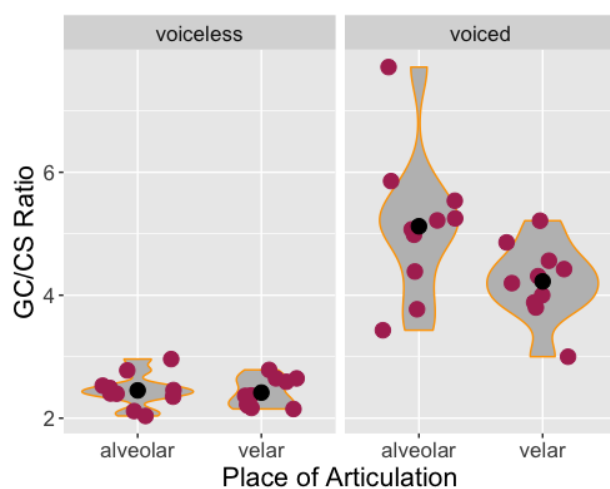


Figure 5.2 Violin plots of ratio of closure duration of singleton and geminate stops for the stimulus speaker.



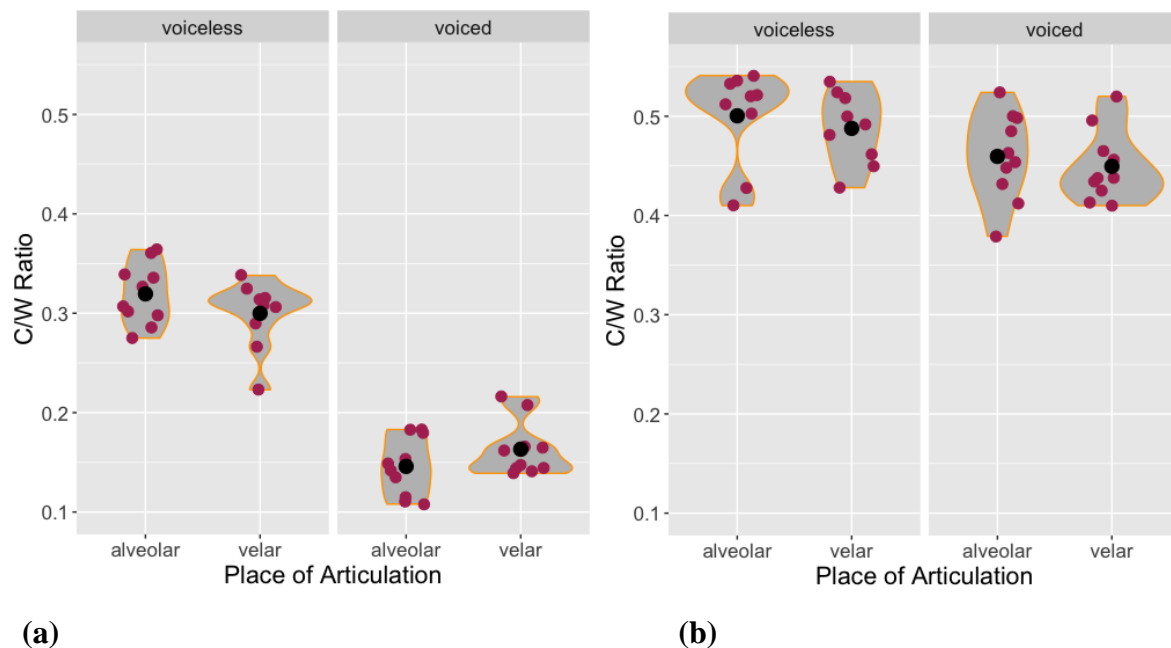


Figure 5.3 Violin plots of ratio of closure (C) to word duration for the stimulus speaker: (a) single consonant (b) geminate consonant.

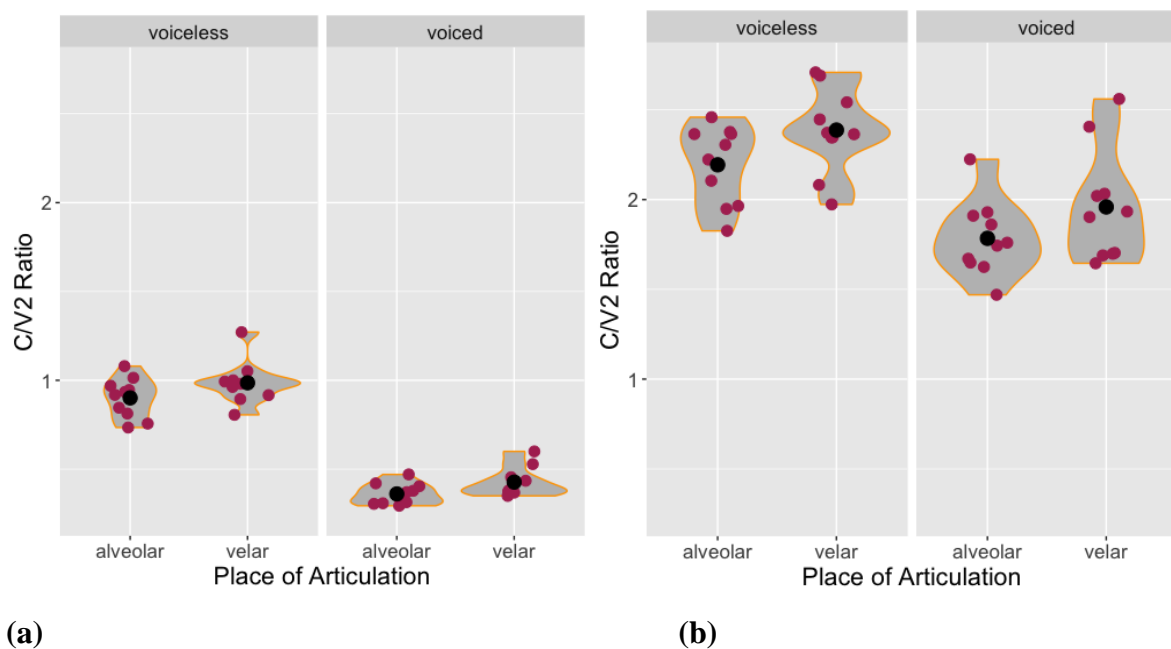


Figure 5.4 Violin plots of ratio of closure to word duration for the stimulus speaker: (a) single consonant (b) geminate consonant.

Mean ratios of geminate to single closure (GC/SC) are about twice for voiced stimuli ([t] 2.45, [d] 5.12, [k] 2.42, [g] 4.22) in Figure 5.4. Therefore, listeners might be able to accurately perceive length contrasts in voiced stimuli.

In terms of geminates, English does not have phonemic level of consonant durational contrast. However, geminate stops occur across the morphological boundary, as in ‘cat tail’.

Studies that used reading passages that included geminate stops across boundary revealed that while an average ratio of [tt]:[t] was 2:1 for American English speakers, the ratio of Japanese was 2.8:1 (Han, 1992). Similar to this finding, Australian English speakers showed comparable results: 1.7:1 (Toda, 2003). Thus, Japanese geminate consonants are much longer than English pseudo-geminates. Note that I used the term ‘duration’ as this section concerns an acoustic property of stimuli. When I use the term ‘length’, I am referring to the related phonological property.

### **5.3.4 Audio clips as stimulus lists**

There were 240 audio clips that were created as experimental stimuli. Each audio clip consisted of a four-word list which contained three fillers and a first of a sound contrast pair, a 300 milliseconds (ms) beep sound and a second of a sound contrast pair. An inter-stimulus interval (henceforth ISI) was 300ms. A beep sound was created by Praat. Each word was concatenated with a 300ms duration between words across all intervals using a Praat script. Regarding a temporal order of words in each word list, a target word (i.e., the first part of a sound contrast pair) never appeared as the final item in the four-word list as shown in Table 5.5. The grey shading indicates contrast pairs in an audio stimulus clip. For example, a participant heard /hapa/, **/detto/**, /ruku/, /gate/ followed by a 300ms beep. Then the counterpart of the sound contrast pair **/deto/** was heard. Thus, the target /detto/ appeared in the second position. This is because two phenomena in memory known as primacy and recency, and also the number of words in the list were considered. While the primacy effect refers to the tendency that the first item in a list is easier to remember/recall, the recency effect refers to the fact that the final item in a list tends to be easier for people to remember (Henry, 2011). In the current study, there were four words in each list. If only middle items were concerned, then this creates a bias towards the second or the third position as the target words. Therefore, although I acknowledge the primacy effect, target words occur except as the final item in a list to minimise the bias. The occurrence of target words at the first position to third position were evenly allocated through all stimuli. Note that fillers whose final syllable is /ku/ or /gu/ never appeared with velar stimuli. Similarly, fillers whose final syllable is /to/ or /do/ never appeared with alveolar stimuli in order to avoid confusion with the target words.

Table 5.5 Example of word lists and position of the target word

Stimulus audio clip	Word lists including a first of the contrast pair					Second of the pair
1	<b>butto</b>	dona	nepu	kede	beep	<b>buto</b>
2	hapa	<b>detto</b>	ruku	gate	beep	<b>deto</b>
3	bona	mihi	<b>ketto</b>	sapu	beep	<b>keto</b>
4	<b>kutto</b>	nozu	ruse	nipa	beep	<b>kuto</b>

From this point, CVCV-structured words are called *singleton*, CVCCV-structured words are called *geminate*, and CVC-structured words are referred to as CVC.

There were 170 clips that contain sound contrast pairs (singleton-geminate, singleton-CVC), i.e., different pairs. Thus, the correct response for these clips is ‘No’. Another 70 clips contained same pairs (e.g., singleton-singleton), whose correct response is ‘Yes’. For the same pairs, physically identical recordings were not used (not the same recording played twice). The number of same pairs presented to the participants were less than that of the number of different pairs in order to reduce the bias towards responding “Yes”. This is based on the assumption that participants expect an even number of ‘Yes’ and ‘No’ responses. The current study would expect that discrimination is difficult, and participants might misjudge a considerable number of different pairs as the same pairs (i.e., ‘Yes’ response). As to the results, the number of each response would be expected to be quite similar, and the number of ‘Yes’ responses would not be hugely greater than the number of ‘No’, as laid out in Table 5.6.

Table 5.6 Total number of contrast pairs in the audio clips for each participant

Response	Alveolar Voiceless	Alveolar Voiced	Velar Voiceless	Velar Voiced	Total
No	Singleton/Geminate	Singleton/Geminate	Singleton/Geminate	Singleton/Geminate	40
No	Geminate/Singleton	Geminate/Singleton	Geminate/Singleton	Geminate/Singleton	40
No	Singleton/CVC	Singleton/CVC	Singleton/CVC	Singleton/CVC	40
No	CVC/Singleton	CVC/Singleton	CVC/Singleton	CVC/Singleton	40
No	Filler/Filler	Filler/Filler	Filler/Filler	Filler/Filler	10
Yes	Singleton/Singleton	Singleton/Singleton	Singleton/Singleton	Singleton/Singleton	20
Yes	Geminate/Geminate	Geminate/Geminate	Geminate/Geminate	Geminate/Geminate	20
Yes	CVC/CVC	CVC/CVC	CVC/CVC	CVC/CVC	20
Yes	Filler/Filler	Filler/Filler	Filler/Filler	Filler/Filler	10
Total	60	60	60	60	240

As a reminder, the main questions in this study were focussed upon whether the learners of Japanese would be able to perceive difference between singleton and geminate stimuli, and whether native Japanese listeners would be able to perceive differences between singleton and CVC stimuli. The hypothesis is that learners of Japanese would more likely show difficulty in discriminating singleton-geminate contrasts. On the other hand, native Japanese listeners would more likely show difficulty when discriminating singleton-CVC contrasts. Additionally, for all groups, the velar contexts would be harder to discriminate than the alveolar contexts. Thus, the expected responses are different for each group.

In order to test the hypothesis, singleton-geminate contrast stimuli for native Japanese listeners served as control for the learners' group because Japanese listeners would clearly hear the difference for their native contrast. Singleton-CVC contrast stimuli for monolingual English listeners served as control for the Japanese group because we assumed monolingual English listeners would have no knowledge of Japanese, in contrast to learners of Japanese. Therefore, they would clearly hear the difference for their native contrast.

### 5.3.5 Auditory memory decision task protocol

Figure 5.5 illustrates the components of the auditory memory decision task protocol. Inter-stimulus is very important as auditory memory remains for a short time (100-500ms), but fades with an increasing interval after 500ms (Massaro, 1974; Pisoni, 1973; van Hesson & Schouten,

1992). In the current study, ISI is at 300ms. Since the target words never appear in the fourth position of each word list, there is at least 1500ms available to listen to a word after the beep. If a target word occurs in the first position, it entails that the participants hold their memory for more than 2000ms. Thus, participants are forced into a phonemic/phonological mode rather than an auditory/acoustic mode. In the meanwhile, inter-trial interval was not fixed because participants needed to click the ‘NEXT’ to hear a next stimulus audio clip by her/himself. This is due to prevent participants from clicking on the same button all the time.

Trial #	Word lists as audio clips									Correct Response
	First	ISI	Second	ISI	Third	ISI	Forth	ISI- Beep-ISI	Another Word	
1	filler	300 ms	<b>geminate</b>	300 ms	filler	300 ms	filler	←900ms→	<b>singleton</b>	No
2	<b>singleton</b>		filler		filler		<b>geminate</b>		No	
⋮	⋮		⋮		⋮		⋮		⋮	
239	filler		<b>CVC</b>		filler		<b>CVC</b>		Yes	
240	filler		filler		<b>singleton</b>		filler		<b>geminate</b>	No

Figure 5.5 Auditory memory decision task protocol.

### 5.3.6 Participants

Three groups of participants took part in the experiment: 24 native Japanese speakers, 28 English-speaking learners of Japanese (i.e., L1 English-L2 Japanese speakers), and 20 monolingual English speakers. There were in total 72 participants and all of them were living in NZ at the time. Ten of them had participated in the previous experiment which was conducted at the end of second semester in 2018. Detail will be shown shortly. All participants were recruited in the middle of the second semester in the 2019 academic year. Participants were recruited via social-media networks, word-of-mouth, and through on-campus recruiting. Participants were informed that the experiment would take approximately 45 minutes in total, for which they received a payment of \$15NZD e-voucher. The recruitment procedures and all the text used were approved by the University of Canterbury Human Ethics (under application number HEC2019/27/LR-PS).

For the learners of Japanese, participants were mainly recruited before/after their Japanese classes by receiving recruitment flyers. All learners were taking or had recently taken a Japanese course as their L2 at universities in NZ when the data was collected. Although ideally, learners of Japanese would be taking Japanese courses at the time, only intermediate level was a whole-year course; other levels were a semester-length course. However, only a few learners had finished their course and most learners were still taking a course.

In this experiment as with the first experiment, all non-native Japanese speakers were native speakers of English, which controls for the effects of participants' first language. Note that although the majority of English speakers for both learners and non-learners (i.e., English monolinguals) groups are NZE speakers, the groups also included two Australian-English speakers and two American-English speakers. Even if participants had some bias related to geminates, these biases should be consistent among non-native speakers of Japanese due to sharing the same language background. Some L2 learners were bilingual, as they spoke heritage languages. For the monolingual group, all participants were university students except for one person. For the Japanese group, nine participants were university students and others were not.

As mentioned at §5.3.4, the hypothesis would be tested by learners of Japanese, and as a control with Japanese listeners for one contrast stimuli and with monolingual English listeners for other contrast stimuli because those stimuli were their native language.

For data analysis, two participants were excluded as they checked the box indicating speech or hearing impairment and three participants were excluded for not matching their information to the selection criteria, leaving a total of 67 participants whose responses were analysed. The number of participants in each group is presented in Table 5.7. Among the participants, six people from the Japanese group, three people from the learners' group, and one person from the non-learners' group had participated in the previous experiment (i.e., well-formedness task in Chapter 4).

Table 5.7 Number of participants in each group for analysis

<b>Participants</b>	<b>Number</b>
Native Speakers of Japanese (i.e., L1 Japanese-L2 English speakers)	24
English-Speaking Learners of Japanese (i.e., L1 English-L2 Japanese speakers)	26
English-Speaking Monolinguals	17
Total	67

### 5.3.7 Procedure

Some explanation of procedure is duplicated from the first experiment as being a basic setup and use of auditory experiment. As in the first experiment, the current experiment was implemented as an online task using an online survey platform called the Speech In Noise 2 (Chan, 2018) via the NZILBB link. Each participant did the task in their own convenient time and place. Participants were asked to read and follow the instructions on screen to perform the

task. Headphones or earphones were required. The entire procedure was described to participants in their native language.

At first, participants completed the audio system test in which they were instructed to listen to the stimuli through headsets. Two audio files, 'dog' and 'book', created by a NZ male speaker, were presented. After participants listened to the audio files one at a time, they typed the word they heard. Then, following instructions on the computer screen, they were asked to complete the auditory memory decision task. All participants were given the following instructions in their native language. "The purpose of this study is to see whether people can memorise spoken words and how recognition memory is different between speakers of different languages." A next screen showed that "For each trial, you will hear a list of 4 words followed by a beep. After the beep, you will hear another word. Your task is to judge whether the word after the beep was in the list of words before the beep. When you click NEXT, you will hear the next word list immediately. Please remember the list and following word play only once." Then, examples were given to participants on the following page. "Task Instruction (Examples), You hear CAP, DOG, MOUSE, SHOES, (beep), FOOT. As FOOT was not in the list before the beep, you click 'No' on the screen. You hear CAP, DOG, MOUSE, SHOES, (beep), DOG. As DOG was in the list before the beep, you click 'Yes' on the screen. The list and another word will play only once. If you are ready, please click FWD twice."

Instruction did not imply anything about what language the words were in. The labels of button such as NEXT, FWD were also in participants' native language.

There was a 120 voiceless word list and 120 voiced word list. All the lists were randomised together and presented in a different random order for each participant, for a total of 240 trials.

The procedure for each trial was as follows, participants heard a list of four words followed by a 300ms beep. After the beep, another word would be heard. After the participants clicked on one of two options 'Yes' or 'No' and a 'NEXT' button on the computer screen, then the next word was heard immediately. For example, a participant heard /hapa/, /detto/, /ruku/, /gate/ followed by a 300ms beep. Then another word /deto/ was heard. As /deto/ was not in the list, the participant should click 'No' on the computer screen, if they can hear the singleton/geminate distinction. The ISI was 300ms. Each audio clip was approximately 6ms. Participants listened to the list and the following word only one time. After finishing the decision task, the participants had to fill out a questionnaire. There was no training phase or practice phase. No feedback was given during the experiment. The whole experiment lasted approximately 40-45 minutes, including reading a consent form, and filling in a questionnaire.

## **5.3.8 Questionnaires and PCA**

### **5.3.8.1 Questionnaires**

The experiment consisted of two parts: auditory memory decision task and filling out questionnaires. The questionnaire contained questions about demographic information, participants' knowledge of Japanese/English and exposure to Japanese/English. These are basically same as those used in the questionnaire of the first experiment (see Appendix D). However, for the questions related to exposure to Japanese, questions that were not effective on the last results were removed. That is, only questions that contributed to the PCA in the first experiment were used. In addition, one question was added to ask whether participants took part in the first experiment.

### **5.3.8.2 Principal component analysis (PCA)**

From this point, those people who participated in the auditory memory decision task will be called listeners. After collecting the responses of the questionnaires, all responses were converted to numeric values in the same manner as the first experiment.

PCA scores for each listener in each group were extracted based on the loadings for each principal component (PC) from the previous experiment (i.e., well-formedness judgment task). There were two components for Japanese group and learners' group in the first experiment. Note that, for English-monolingual listeners, PCA scores were not extracted. This is because the purpose of the current experiment is simply to test their perception of non-native sounds in order to compare with that of L1 English-L2 Japanese listeners.

As a reminder, for Japanese group, the two principal components (factors) were revealed by PCA in the first experiment, as shown in Table 5.8. The first component (PC1) was related to exposure to English in NZ. More specifically, the questions in PC1 were about self-reported levels of proficiency in English and period of stay in NZ. The second component (PC2) was related to exposure to other languages rather than English and overall knowledge of languages. Thus, considering the questions of the two components, PC1 was labelled as 'PC-English' and PC2 was labelled as 'PC-Language'.



Table 5.8 Two principal components used in the first experiment for native Japanese listeners

(PC1)	How well can you understand/read English?
PC-English	Approximately how many years have you been in New Zealand?
	How well can you speak English?
	Languages in which you can have an everyday conversation
(PC2)	If you are studying Linguistics, at what level?
PC-Language	Languages in which you can have an everyday conversation

For learners, the two principal components were also revealed by PCA in the first experiment, as shown in Table 5.9. From the questions, it can be seen that the first component (PC1) was related to exposure to Japanese in Japan. Thus, PC1 was labelled as ‘PC-JJ’. The second component (PC2) was related to exposure to Japanese in NZ. Therefore, PC2 was labelled as ‘PC-JNZ’.

Table 5.9 Two principal components used in the first experiment for L1 English-L2 Japanese listeners

(PC1)	How long have you been in Japan?
PC-JJ	What was the purpose of the visit?
“exposure to Japanese in Japan”	How many times have you been to Japan?
	When did you last visit Japan?
(PC2)	How often do you access websites about modern Japanese culture?
PC-JNZ	How often do you browse Japanese written cartoons or magazines?
“exposure to Japanese in NZ”	How often do you watch Japanese TV programs or Japanese movies on the internet (with English subtitles)?

Based on the principal components above, PC scores for the current experiment were extracted by going through the following procedure. Firstly, original PCA was performed on a matrix of the first experiment. These factor loadings together with each participant’s Likert data on the questionnaire were used to predict PC scores for the present experiment, as shown in Table 5.10. These principal components will be explored in the statistical analyses in the present experiment.

Table 5.10 Two factor scores revealed by PCA for L1 English-L2 Japanese listeners

Participant	PC1	PC2
1	0.28340611	0.40500367
2	0.28429944	-0.33771113
3	-0.26428327	-0.26523911
...	...	...
22	-0.01410027	-0.56509682
23	1.97510532	0.18237317

### 5.3.9 Statistical analysis

#### 5.3.9.1 Factors

Trials were coded as “1” when participants responded correctly in judging the stimuli given, and “0” when they responded incorrectly. The shorthand ‘correct’ is used to indicate predicted answers if all phonological distinctions, in both languages, were heard perfectly. The dependent variable was the test trial outcome (binary: “1” correct or “0” incorrect) in the experiment (i.e., accuracy). There were two broad categories for the independent factors: phonological and extra-linguistics factors based on the first experiment. Independent factors on responses are summarised in Table 5.11. Independent variables as phonological factors are voicing types (2 levels: voiced/voiceless) and POA (2 levels: alveolar/velar). Independent variables as extra-linguistic factors are PCA scores, stimulus pair (2 levels: same/different) and number of trials. For stimulus pair, if given stimuli A and B are being successfully discriminated, four possible sequences are observed. For geminate discrimination, <singleton-singleton> <geminate-geminate> are the same pair, and <singleton-geminate> <geminate-singleton> are the different pair. For CVC discrimination <singleton-singleton> <CVC-CVC> are the same pair and <singleton-CVC> <CVC-singleton> are the different pair. Number of trials (1-240) was standardised by converting the variable to a z-score.

For analysing data from discrimination tasks, a perceptual sensitivity measure such as  $d'$  is commonly used for assessing listeners’ performance instead of percent correct measures. The  $d'$  analysis is on the basis of signal-detection theory (Macmillan & Creelman, 2005), measuring how easily listeners can detect the presence of the signal (i.e., targets in different trials). In addition, “a signal detection analysis should provide us with a clear separation between sensitivity and bias” (Gerrits & Schouten, 2004, p. 365). However, I did not choose to use  $d'$ . This is because in the current study, the first stimulus words appeared in one of the three positions, rather than staying fixed. That is, they never appeared in the same

position of each word list. Therefore, participants' sensitivity towards to the contrasts might differ among the three positions. However, calculation of original  $d'$  scores does not consider such rotations. Therefore, calculating  $d'$  is computationally complex. In terms of response bias, as discussed in §5.3.4, this study considered participants' bias towards responding "Yes" (i.e., "same"), since it was expected that the current discrimination task would be difficult. Thus, the number of same pairs were reduced by half, which would help to remove the undesirable response bias in the percent correct measurement. For those reasons, I did not use  $d'$  for the dependent variable for the current data analysis.

Note that, although there are three groups in the full dataset, data for each group would be submitted to a separate mixed effect model. Therefore, in the current analyses *Group* is not a factor.

Table 5.11 Factors considered in analysis of dataset

	<b>Factors</b>	<b>Levels</b>
Phonological Factors	VOICING TYPE	voiced/ voiceless
	POA	alveolar/ velar
	Pair	same/different
Extra- Linguistic Factors	PC-JJ (Exposure to Japanese in Japan)	Index score
	PC-JNZ (Exposure to Japanese in NZ)	Index score
	PC-English (Exposure to English in NZ)	Index score
	PC-Language (Exposure to other languages in NZ)	Index score
other	zTrial (z-scored 1-240 trial numbers)	Index score

### 5.3.9.1.1 Phonological factors

#### *Voicing type*

There were two voicing types, voiced and voiceless stops. If acoustic signals are influential factors for discriminating non-native segmental length contrasts, voiced stimuli should be perceived well compared to voiceless stimuli. Mean ratios of geminate to single closure (GC/SC) are about twice for voiced stimuli ([t] 2.45, [d] 5.12, [k] 2.42, [g] 4.22). Therefore, voiced stimuli might be perceived better than voiceless stimuli.

#### *POA*

In this experiment, when the second consonants in the words are the velar stops /k/, /g/, the following vowel is /u/ (e.g., /keku/, /geku/, /segu/, /migu/). They are called velar stimuli. When

the second consonants in the words are the alveolar stops /t/, /d/, the following vowel is always /o/ (e.g., /kuto/, /keto/, /sudo/, /nedo/). They are called alveolar stimuli. Thus, POA is the factor distinguishing these stimuli types. As a reminder, the prediction is that the velar stimuli would be discriminated less well compared to the alveolar stimuli as the velars are acoustically less salient.

### 5.3.9.1.2 Extra-linguistic factors (PCA)

I assume that rating values for pronunciation types reflects each L1 English-L2 Japanese listeners' exposure levels to Japanese language. Similarly, L1 Japanese-L2 English listeners' exposure levels to English language will influence rating values for the non-native sound contrasts.

### 5.3.9.2 Mixed effects models

All responses in the experiment for each group were analysed using the *lme4* package in R (Bates et al., 2015; R Core Team, 2018) to perform a mixed effects logistic regression model with the *glmer()* and *bobyqa* optimizer. In order to measure performance across participants for each group, firstly distribution of entered responses (yes/no) in the memory decision task was checked by using a histogram. Listeners that lie outside of the overall pattern of distribution were considered. Two Japanese listeners in Japanese group, and three L1 English-L2 Japanese listeners were removed. Therefore, there were 62 listeners in total in this final stage. Secondly, mean reaction time in the perception task was analysed. This is the same procedure of the first experiment. There were 14,880 responses in total. Then, the distribution of the data was checked and reaction times more than two standard deviations above the mean were removed as outliers from further analysis. Therefore, in the current research, 69 data points (0.46 %) of observations were removed, and 14,811 observations remained for analysing the responses shown in Table 5.12

Table 5.12 Numbers of stimuli of each place of articulation by language group

Group	# of listeners	Items			Total of Items
		Alveolar	Velar	Fillers	
Native Japanese listeners	22	2,407	2,414	439	5,260
L1 English-L2Japanese listeners	23	2,513	2,509	452	5,474
Monolingual English listeners	17	1,870	1,867	340	4,077
Total	62	6,790	6,790	1,231	14,811

As in the first experiment, the statistical models that predict the responses in the current study were created the following manner. First, a binomial model with a fully crossed and specified random effects structure was created (Jaeger, 2011). The dependent variable is correct versus incorrect trials (i.e., accuracy). Three linguistic predictors (VOICING TYPE, POA, PAIR), two extra-linguistic predictors (PC1, PC2) and a trial number (zTrial) were included in this model initially. Two four-way interactions of (VOICING TYPE, POA, PAIR, PC1), and (VOICING TYPE, POA, PAIR, PC2) were tested and zTrial was a fixed effect at the outset. All slopes were initially included. Then, if the inclusion of a slope led to convergence errors, the slope that contributes least to the model is dropped in order to obtain convergence. Using stepwise regression, models are compared to each other using an `anova()` function to see which fits the best, based on the Akaike Information Criterion (AIC) and considering  $p$  values (Baayen et al., 2008; Jaeger, 2008). If the ANOVA test shows a significant ( $p < .05$ ) improvement in model fit, the factor is retained. Non-significant factors are removed from the model one-by-one. Non-significant main effects are also removed from the model as well (Barr et al., 2013). Estimated values in all results are given in log odds. Plots from the best-fitting models were made using the `ggPredict` function in the `ggeffects` package in R (Lüdtke, 2018).

Note that for the current analyses, each model was created for each language group by performing subset data analyses for singleton-geminate contrast and singleton-CVC contrast, respectively. That is, for discrimination of singleton-geminate contrast, the dataset contains <singleton-singleton> <geminate-geminate> as the same pair, and <singleton-geminate> <geminate-singleton> as the different pair. For discrimination of singleton-CVC contrast, the dataset contains <singleton-singleton> <CVC-CVC> as the same pair and <singleton-CVC> <CVC-singleton> as the different pair.

As with the first experiment, a Variance Inflation Factor (VIF) score was calculated for each factor used in the best models, and all VIF scores were less than 5, which suggests the model has no multicollinearity.

## 5.4 Results

### 5.4.1 Singleton-Geminate contrasts

This section considers the results of singleton-geminate contrast (7,408 tokens, 80 words from 62 listeners) with relevant predictions. We begin by looking at the results of native Japanese listeners serving as a control group since the consonant length contrast is phonemic and lexical in Japanese.

#### 5.4.1.1 Native Japanese listeners

Firstly, in order to analyse the results of native sound contrasts (i.e., singleton-geminate contrast), the data (2,630 tokens) were fitted in the mixed effect logistic regression model. Three linguistic predictors (VOICING TYPE, POA, PAIR), two extra-linguistic predictors (PC-English, PC-Language) and a trial number (zTrial) were included in this model initially. The interactions of VOICING TYPE, POA, PAIR, PC-English, and VOICING TYPE, POA, PAIR, PC-Language were tested and zTrial was as a fixed effect at the outset, and dimensions that were not significant were removed one-by-one in a stepwise manner from the model. The best-fit model has an interaction PAIR and PC-English with fixed effect as zTrial. Along with these factors, SUBJECT and WORD are included as a random intercept in the model, as well as a random slope of zTrial by subject. Thus, there is no effect of VOICING TYPE, POA and PC-Language. The below is the best model predicting the response of this group.

```
glmer (correct ~ PAIR * PC_English + zTrial+ (1+zTrial|subject)
+ (1|word), data = geminate, family = "binomial")
```

The results of the model are presented in Table 5.13. The reference category is *same pair*. The positive effect of the trial is detected ( $\beta = 0.26$ ,  $z=3.07$ ,  $p < 0.01$ ), suggesting discrimination accuracy increased during the course of trials in the experiment. There is also a significant interaction of PAIR with PC\_English ( $\beta = -0.40$ ,  $z = -2.89$ ,  $p < 0.01$ ). This indicates that discrimination accuracy depends on the relation between pairs of stimuli and the extra-linguistic factor, PC-English. This interaction is illustrated in Figure 5.6. The figure on the left shows the relationship of accuracy and types of stimuli. The x-axes are the four consonant types in C2 of stimuli, and the y-axis denotes the percentage of correct trials. The black dot indicates the mean value in the figures. Wider sections of each plot represent which values in the sections have higher frequency while the thinner sections represent lower frequencies. The same description applies hereafter for all of the violin plots. As shown in the figures, when stimuli of different pairs are presented, rate of correct trials are slightly lower than that of same pairs. However, overall listeners discriminate different pairs well, which confirms **Prediction 1a (1)**. Geminate stimuli are phonemic/phonological categories for native Japanese listeners, so that they demonstrate phonemic perception of discrimination between the two stimuli. Regarding **Prediction 1b**, mediating effects of phonetic salience were not found.

The figure on the right shows the probability predicted by the best-fitted model. The x-axes present individuals' PC-English scores. A higher PC-English indicates that listeners have

more exposure to English than listeners with lower PC-English scores. The y-axis is the predicted accuracy. The predicted plots suggest that Japanese listeners with higher exposure to English are more likely to judge correctly same pairs than Japanese listeners with lower exposure to English. On the other hand, participants with higher exposure to English are more likely to have difficulty perceiving the native-sound contrasts than listeners who have had less exposure to English. Surprisingly, this seems to suggest that exposure to English interferes with the ability of native speakers of Japanese to accurately attend to the singleton/geminate contrast.

Table 5.13 Model summary for singleton-geminate discrimination by native Japanese listeners

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	1.96215	0.2293	8.557	< 2e-16
pair different	-0.58876	0.1185	-4.969	6.75E-07
PC_English	0.14155	0.25056	0.565	0.57213
zTrial	0.26319	0.08558	3.075	0.0021
pair different: PC_English	-0.40173	0.1387	-2.896	0.00377

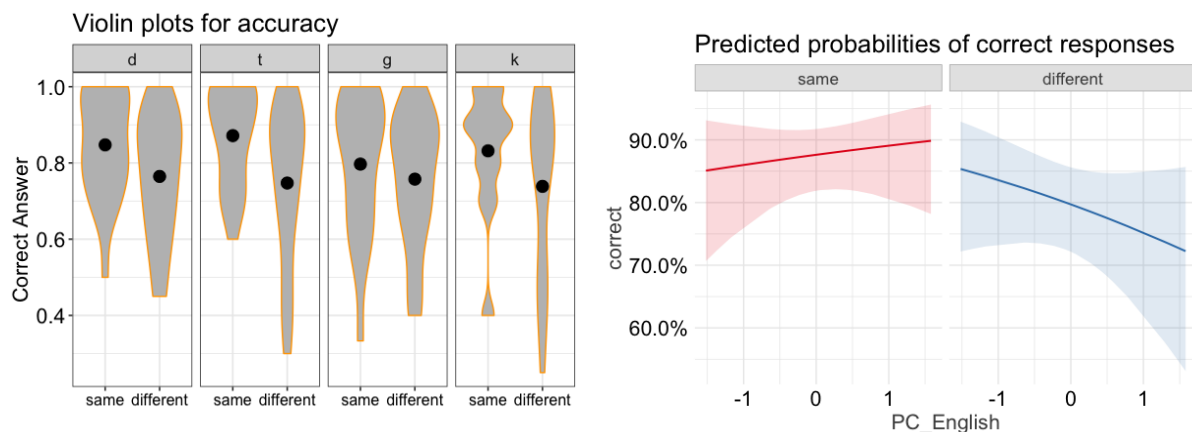


Figure 5.6 Distribution and probability density of discrimination accuracy on singleton-geminate contrast by native Japanese listeners for the raw data (left) and predicted interaction between stimulus pair and PC-English (right).

#### 5.4.1.2 L1 English-L2 Japanese listeners (learners of Japanese)

Next, focusing on discrimination accuracy of L1 English-L2 Japanese listeners was examined, using subset data (2,739 tokens). The process to obtain the best fit model is the same as that of native contrasts described before. The interactions of VOICING TYPE, POA, PAIR, PC-JJ, and VOICING TYPE, POA, PAIR, PC-JNZ were tested and zTrial was as a fixed effect at the

outset, and dimensions that were not significant were removed one-by-one in a stepwise manner from the model. The best-fit model has two-way interactions between POA and PAIR, VOICING TYPE and PAIR, PAIR and PC-JJ, and PAIR and PC-JNZ. In addition, SUBJECT and WORD are included as a random intercept. The below is the best model predicting the response of this group.

```
glmer (correct ~ POA * PAIR + VOICING_TYPE_ * PAIR + PAIR * PC-JJ
+ PAIR * PC-JNZ +(1|subject) + (1|word), data = geminate, family = "binomial")
```

The results of the model are presented in Table 5.14. The reference categories are *voiced alveolar same pair*. There is no effect of zTrial, indicating the accuracy did not improve as the number of trials increased. There are four significant two-way interactions, between POA and PAIR ( $\beta = -0.56, z = -2.90, p < 0.01$ ), PAIR and VOICING TYPE ( $\beta = -0.88, z = -4.56, p < 0.001$ ), PAIR and PC-JJ ( $\beta = -0.56, z = -5.73, p < 0.001$ ), PAIR and PC-JNZ ( $\beta = -0.66, z = -6.22, p < 0.001$ ), respectively. Thus, four significant interactions of *pair* with other variables are detected, but no three-way interaction with PAIR is found.

The interaction of POA and PAIR indicates that discrimination accuracy depends on the relation between place of articulation and type of stimulus pair. The effect of POA is significantly negative on velars in different pairs, suggesting that discrimination of velar stimuli in different pairs is difficult for the listeners regardless of voicing types. The interaction of VOICING TYPE and PAIR indicates that discrimination accuracy also depends on the relation between place of articulation and voicing type in pairs. The effect of VOICING TYPE is significantly negative for voiceless different pair, suggesting that discrimination of voiceless stimuli in different pairs is difficult for the listeners. They answer ‘same’ more than ‘different’ in the voiceless velar context – though not nearly to the same extent as the “same” token. These interactions are illustrated in Figure 5.7.

The figure on the left shows the relationship of accuracy and types of stimuli, and this figure confirms **Prediction 1a (2)** that some listeners show a sensitivity to phonemic distinctions between geminate stimuli, although the stimuli are non-native language phonological categories. The figure on the right shows the probability predicted by the best-fitted model. The subset analysis was performed, and it was found that when stimuli of same pairs are presented, the effect of POA ( $\beta = 0.007, z = 0.27, p = 0.78$ ), and the effect of VOICING TYPE ( $\beta = 0.26, z = 0.9, p = 0.36$ ) are not significant. On the other hand, when stimuli of different pairs are presented, the effect of POA is significant ( $\beta = -0.47, z = -3.07, p < 0.01$ ). Thus,



discrimination of velar stimuli is more difficult than that of alveolar stimuli, which confirms **Prediction 1b (1)**. In addition to that, the effect of VOICING TYPE is also significant ( $\beta = -0.70$ ,  $z = -4.91$ ,  $p < 0.001$ ). Thus, discrimination of voiceless stimuli is more difficult than that of voiced stimuli, which confirms **Prediction 1b (2)** regarding to mediating effects of phonetic salience.

Table 5.14 Model Summary for Singleton-Geminates discrimination by L1 English-L2 Japanese listeners

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	0.8623	0.20596	4.187	2.83E-05
POA velar	0.11378	0.17649	0.645	0.51916
Pair different	-0.05092	0.16753	-0.304	0.76119
voicing_type voiceless	0.21806	0.17613	1.238	0.21569
PC-JJ	0.45327	0.16878	2.686	0.00724
PC-JNZ	0.55807	0.17593	3.172	0.00151
POA velar: pair different	-0.56275	0.1938	-2.904	0.00369
Pair different: voicing_type voiceless	-0.88209	0.19325	-4.565	5.01E-06
Pair different: PC-JJ	-0.56886	0.09926	-5.731	9.98E-09
Pair different: PC-JNZ	-0.66286	0.10652	-6.223	4.88E-10

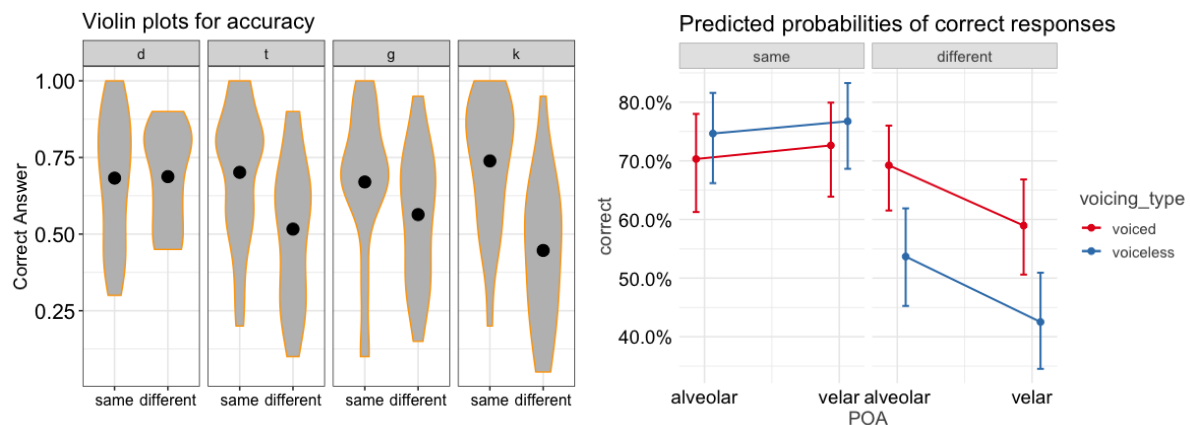


Figure 5.7 Distribution and probability density of discrimination accuracy on singleton-geminate contrast by L1 English-L2 Japanese listeners for the raw data (left) and predicted interaction between POA, voicing type and stimulus pair.

For the extra-linguistics factors, two two-way interactions PAIR and PC-JJ ( $\beta = -0.56$ ,  $z = -5.73$ ,  $p < 0.001$ ), PAIR and PC-JNZ ( $\beta = -0.66$ ,  $z = -6.22$ ,  $p < 0.001$ ) are detected, as shown in Figure 5.8, and these interactions are related to **Prediction 1c** regarding the effects of the degree of exposure of the individuals to the target language. The x-axis presents individuals' PC-JJ scores. The y-axis is the predicted accuracy. The interaction between pair and PC-JJ

(Figure 5.8: left) suggests that the higher the PC-JJ scores are, the better listeners perceive same pairs. On the other hand, the higher the PC-JJ scores are, the less listeners discriminate different pairs. The effect of PC-JNZ (Figure 5.8: right) shows the same trend. These results suggest that participants with high exposure to Japanese either in Japan or in NZ detect same pairs well, but do not discriminate well for different pairs in non-native contrasts. As opposed to **Prediction 1c**, greater exposure to a language presumably does not lead to an increased ability to identify the relevant phonological contrasts. Thus, for the different pairs, the performance of L1 English-L2 Japanese listeners with high exposure to Japanese is not better than that of L1 English-L2 Japanese listeners with low exposure to Japanese.

Note it is interesting that listeners would respond ‘different’ so much to the ‘same’ pairs. This is quite noticeable in comparison to responses of Japanese group in the same pairs. We can speculate that it might reflect that listeners notice some acoustic length difference might be important, and they can hear that the tokens given to them are not identical.

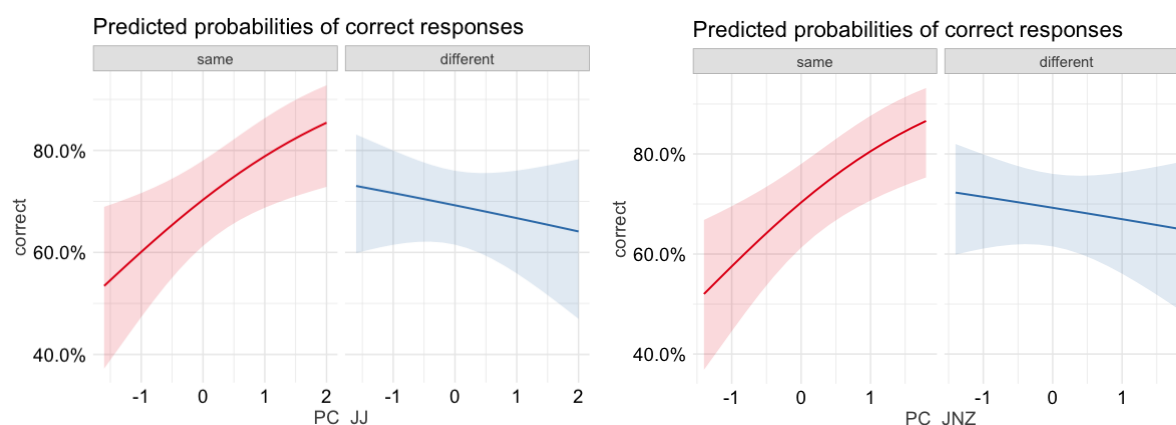


Figure 5.8 Interaction between pair and PC-JJ (left), and between pair and PC-JNZ (right)

#### 5.4.1.3 Monolingual English listeners

This section considers the results of monolingual English listeners (2,039 tokens from 17 listeners). The data (2,039 responses) were fitted in the mixed effect logistic regression model. The process to obtain the best fit model is similar as the two other groups, except for extra-linguistic factors. As mentioned in the previous section, this group does not have PC scores. The dependent variable is correct trials (i.e., accuracy). There were three linguistic predictors (VOICING TYPE, POA, PAIR) and a trial number (zTrial) included in this model initially. The interactions of VOICING TYPE, POA, and PAIR were tested and zTrial was as a fixed effect at the outset. Dimensions that were not significant were removed one-by-one in a

stepwise manner from the model. The best-fit model has a two-way interaction of VOICING TYPE and PAIR. Along with these factors, SUBJECT and WORD are added as a random intercept in the model. The following is the best model predicting the response of this group.

```
glmer (correct ~ VOICING TYPE * PAIR + (1|subject) + (1|word),  
data = geminate, family = "binomial")
```

The results of the model are presented in Table 5.15. The reference category is *voiced same pair*. A significant two-way interaction between VOICING TYPE and PAIR is detected ( $\beta = -0.56$ ,  $z = -2.39$ ,  $p < 0.05$ ). This indicates that discrimination accuracy depends on the relation between voicing type and type of stimulus pair. The effect of VOICING TYPE is strong in different pairs negatively, especially for voiceless stimuli, suggesting that discrimination of voiceless stimuli in different pairs is difficult for the listeners. This is illustrated in Figure 5.9. The figure on the left shows the relationship of accuracy and types of stimuli. The x-axes are the four consonant types and the y-axis denotes the percentage of correct trials. Listeners tend to answer “same” more than “different” for different pairs. The figure on the right shows the probability predicted by the best-fitted model. As shown in these figures, when stimuli of same pairs are presented, there seems to be no significant difference regardless of voicing types. On the other hand, when stimuli of different pairs are presented, the accuracy is different between voiced and voiceless stimuli. The subset analysis was performed, and it was found that the effect of VOICING TYPE is significant when stimuli of different pairs are presented ( $\beta = -0.51$ ,  $z = -2.71$ ,  $p < 0.01$ ). On the other hand, when stimuli of same pairs are presented, the effect of VOICING TYPE is not found ( $\beta = 0.18$ ,  $z = 0.56$ ,  $p = 0.57$ ). Thus, discrimination accuracy of voiceless stimuli in different pairs is significantly lower than that of voiced stimuli in the different pairs. This confirms **Prediction 1b (2)** that listeners are quite sensitive to acoustic signals, and contrasts in voiced stimuli are more likely to be discriminated better than contrasts in voiceless stimuli. This is based on the degree of acoustic difference in the voiced and voiceless stimuli; mean ratios of geminate to single closure (GC/SC) are about twice for voiced stimuli.

Table 5.15 Model Summary for singleton-geminate discrimination by monolingual English listeners

	Estimate	Std.Error	z value	Pr(> z )
(Intercept)	1.3981	0.1779	7.86	3.84E-15
voicing_type voiceless	0.1197	0.2163	0.554	0.5799
pair different	-1.7931	0.1654	-10.838	< 2e-16
voicing_type voiceless: pair different	-0.566	0.236	-2.399	0.0165

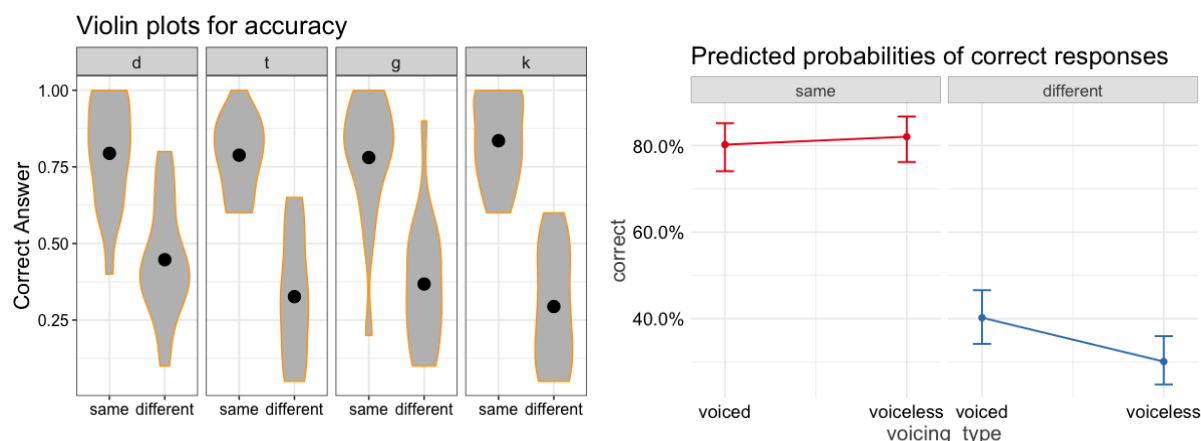


Figure 5.9 Distribution and probability density of discrimination accuracy on singleton-geminate contrast by monolingual English listeners for the raw data (left) and predicted interaction between voicing type and stimulus pair (right).

Overall, results showed that monolingual English listeners answered “same” more than “different” for different pairs. That is, while they answered “same” 80% of the time for same stimuli and 60-70% for “different” stimuli, this indicated that in both cases the tendency to hear them is the same. But it is strong for the actual same stimuli. Thus, discrimination of non-native sound contrasts is difficult for the naïve listeners, which confirms **Prediction 1a (3)**. Monolingual English listeners do not discriminate between consonants varying in duration as the listeners do not have phonemic categories according to consonant length. This clearly reveals that the L2 learners are using some additional knowledge beyond the monolinguals.

Interestingly, the discrimination accuracy of the same stimuli by this group (Figure 5.9, right) is higher than that of the learners’ group. This may indicate that L2 learners entertain the idea that similar sounds might not be the ‘same’ in Japanese.

## 5.4.2 Results of Singleton-CVC contrasts

This section considers the results of singleton-CVC contrasts (7,407 tokens from 62 listeners). We begin by looking at the results of monolingual English listeners who serve as a control group in this context.

### 5.4.2.1 Monolingual English listeners

This section considers the results of monolingual English listeners. In order to analyse the results of native sound contrasts (i.e., CVC and singleton), subset data (2,037 tokens) were fitted in the mixed effect logistic regression model. Three linguistic predictors (VOICING TYPE, POA, PAIR) and a trial number (zTrial) were included in this model initially. The interactions of VOICING TYPE, POA, and PAIR were tested and zTrial was as a fixed effect at the outset, and dimensions that were not significant were removed one-by-one in a stepwise manner from the model. The best-fit model has a two-way interaction of POA and PAIR with SUBJECT and WORD as a random intercept. The below is the best model predicting the discrimination accuracy of native contrasts for this group.

```
glmer (correct ~ POA * PAIR + (1|subject) + (1|word),  
data = CVC, family = "binomial")
```

The results of the model are presented in Table 5.16. The reference category is *alveolar same pair*. There are no fixed effects of POA, PAIR, VOICING TYPE and zTrial, however, a significant two-way interaction between POA and PAIR resulted ( $\beta = -0.63$ ,  $z = -2.52$ ,  $p < 0.05$ ). This indicates that discrimination accuracy depends on the relation between place of articulation and type of stimulus pair. The effect of POA is significantly negative for different velar pairs, suggesting that discrimination of velar stimuli in different pairs is difficult for the listeners. This is illustrated in Figure 5.10. The figure on the left shows the relationship of accuracy and types of stimuli. The figure on the right shows the probability predicted by the best-fitted model and confirmed **Prediction 2a (1)**. Monolingual English listeners demonstrate phonemic perception of CVC stimuli that are within their L1 phonological categories with very good perceptual differentiation of contrasts. The subset analysis was performed, and it was found that the effect of POA is significant when stimuli of different pairs are presented ( $\beta = -0.71$ ,  $z = -4.83$ ,  $p < 0.001$ ). On the other hand, when stimuli of same pairs are presented, the effect of POA is not found ( $\beta = -0.02$ ,  $z = -0.08$ ,  $p = 0.93$ ). Thus, the results confirm **Prediction 2b** regarding mediating effects of phonetic salience. When listeners show difficulty perceiving

contrastive pairs, singleton and geminate /k/, /g/ followed by /u/ (i.e., velar stimuli) are discriminated less well compared to singleton and geminate /t/, /d/ followed by /o/, because the velar context is less phonetically salient than the alveolar contexts. The low accuracy in the velar context might suggest that the vowel [u] is not salient and very vowel-like as we discussed in §5.2.3.

Table 5.16 Model summary for singleton-CVC discrimination by monolingual English listeners

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	1.73762	0.21924	7.925	2.27E-15
POA velar	-0.04717	0.21439	-0.22	0.8258
Pair different	-0.12777	0.18145	-0.704	0.4813
POA velar: pair different	-0.63155	0.25044	-2.522	0.0117

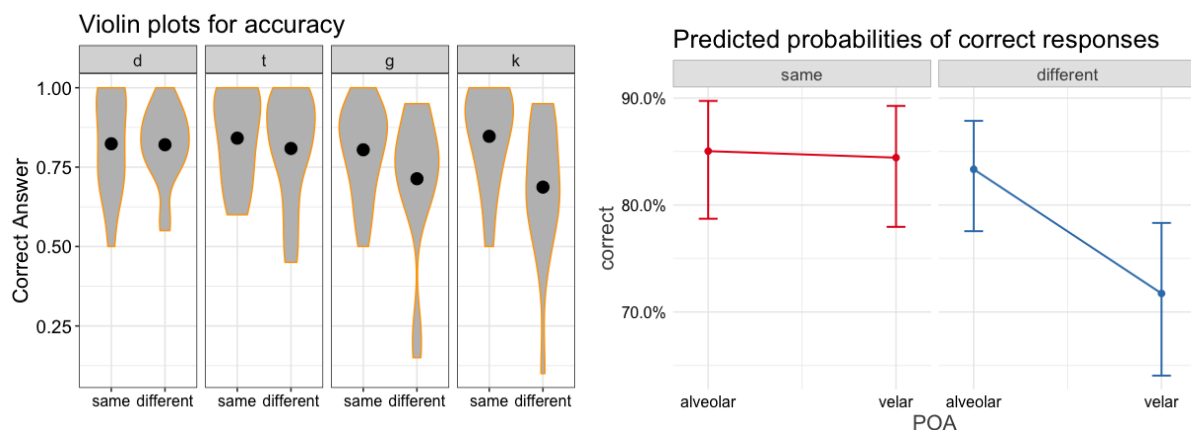


Figure 5.10 Distribution and probability density of discrimination accuracy on singleton–CVC contrast by monolingual English listeners for the raw data (left) and predicted interaction between POA and stimulus pair.

#### 5.4.2.2 L1 English-L2 Japanese listeners (learners of Japanese)

This section considers the results of learners of Japanese. In order to analyse the results of native sound contrasts (i.e., singleton and CVC contrasts), subset data (2740 tokens) were fitted in the mixed effect logistic regression model. There were three linguistic predictors (VOICING TYPE, POA, PAIR), two extra-linguistic predictors (PC-JJ, PC-JNZ) and a trial number (zTrial) were included in this model initially. The interactions of VOICING TYPE, POA, PAIR, PC-JJ, and VOICING TYPE, POA, PAIR, PC-JNZ were tested and zTrial was as a fixed effect at the outset, and dimensions that were not significant were removed one by one in a stepwise manner from the model. The best-fit model has a three-way interaction between POA, PAIR and PC-JJ. In addition to the three-way interaction, the model has a two-way interaction

between and PAIR and PC-JNZ. SUBJECT and WORD are included as a random intercept, and a random slope of zTrial by subject in the model. The below is the best model predicting the response of this group.

```
glmer (correct ~ POA * PAIR * PC-JJ + PAIR * PC-JNZ + zTrial
+ (1+zTrial|subject) + (1|word), data = CVC, family = "binomial")
```

The results of the model are presented in Table 5.17. The reference categories are *alveolar same pair*. A positive effect of zTrial was detected ( $\beta = 0.15$ ,  $z = 2.41$ ,  $p < 0.05$ ), suggesting discrimination accuracy increased each trial during the course of the experiment. There is also a three-way interaction between POA, PAIR and PC-JJ ( $\beta = -0.69$ ,  $z = -3.05$ ,  $p < 0.01$ ). This indicates that discrimination accuracy depends on the relation between place of articulation, pairs of stimuli and the extra-linguistic factor, PC-JJ (i.e., exposure to Japanese in JP). In addition to this three-way interaction, there is also a significant two-way interaction of PAIR with PC-JNZ ( $\beta = -0.76$ ,  $z = -6.34$ ,  $p < 0.001$ ). These results indicate that the accuracy of discrimination also depends on the relation between stimulus pair and the extra-linguistic factor, PC-JNZ (i.e., exposure to Japanese in NZ). These interactions are illustrated in Figure 5.11 and in Figure 5.12.

Firstly, Figure 5.11 shows the relationship of accuracy and types of stimuli. As shown in the figure, it seems there is no difference regardless of place of articulation or voicing types in the same pairs. However, when stimuli of different pairs are presented, rate of correct trials are lower for velar than that of alveolars. That is, geminate /k, g/ followed by /u/ is slightly more difficult for listeners to discriminate.

Table 5.17 Model summary for singleton-CVC discrimination by L1 English-L2 Japanese listeners

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	1.40826	0.24844	5.669	1.44E-08
POA velar	0.1772	0.20829	0.851	0.3949
pair different	0.20551	0.15878	1.294	0.19556
PC-JJ	0.31183	0.24775	1.259	0.20817
PC-JNZ	0.93146	0.23526	3.959	7.52E-05
zTrial	0.1516	0.06287	2.411	0.0159
POA velar: pair different	-0.97405	0.22188	-4.39	1.13E-05

POA velar: PC-JJ	0.33212	0.1869	1.777	0.07557
pair different: PC-JJ	-0.10666	0.16119	-0.662	0.50818
pair different: PC-JNZ	-0.76288	0.12017	-6.348	2.18E-10
POA velar: pair different: PC-JJ	-0.69218	0.22672	-3.053	0.00227

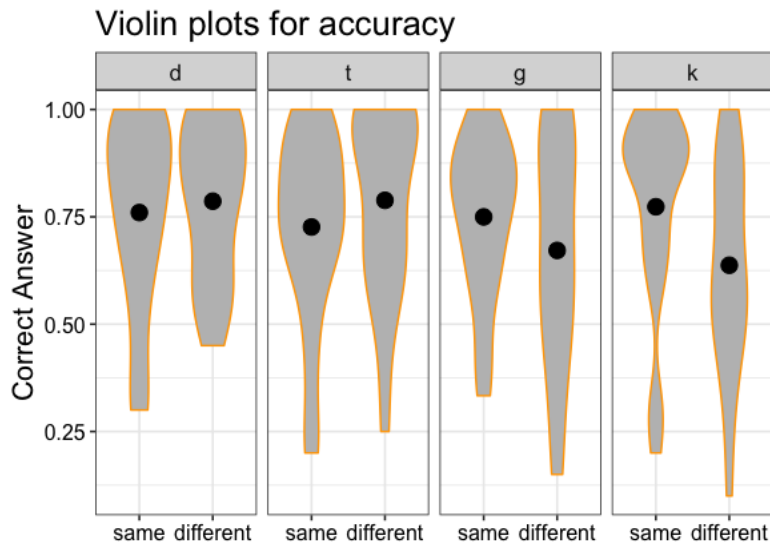


Figure 5.11 Discrimination and probability density of discrimination accuracy on singleton-CVC by L1 English-L2 Japanese listeners for the raw data.

Next, in Figure 5.12 (left), the effect of PC-JJ (i.e., exposure to Japanese in Japan) is observed by interaction with place of articulation and stimuli pairs. The x-axes present individuals' PC-JJ scores. A higher PC-JJ indicates that listeners have more exposure to Japanese in Japan than listeners with lower PC-JJ scores. The y-axis is the predicted discrimination accuracy. When stimuli of same pairs are presented, listeners with higher exposure to Japanese in Japan are more likely to discriminate pairs better than listeners with lower PC-JJ, regardless of place of articulation. On the other hand, when stimuli of different pairs are presented, native contrasts for velar are rather difficult for listeners with high PC-JJ to discriminate, in comparison to that for alveolar. This suggests that exposure to Japanese in Japan influences discrimination of native sound contrasts negatively. This confirms **Prediction 2b** regarding mediating effects of phonetic salience and **Prediction 2c (1)** regarding individual's language experience. The performance of L1 English-L2 Japanese listeners with high exposure to Japanese are lower than that of L1 English-L2 Japanese listeners with low exposure to Japanese in the velar context which is acoustically less salient than the alveolar context. On the other hand, the effect of PC-JNZ (i.e., exposure to Japanese in NZ) works differently in Figure 5.12 (right plots). The x-axes present individuals' PC-JNZ scores. A



higher PC-JNZ indicates that listeners have more exposure to Japanese in NZ than listeners with lower PC-JNZ scores. The y-axis is the predicted accuracy. Listeners with high PC-JNZ are more likely to perceive same pairs better than listeners with low PC-JNZ. For different pairs, PC-JNZ has a little effect. Overall, the higher the PC-JNZ scores are, the better their discrimination accuracy on singleton-CVC contrasts is.

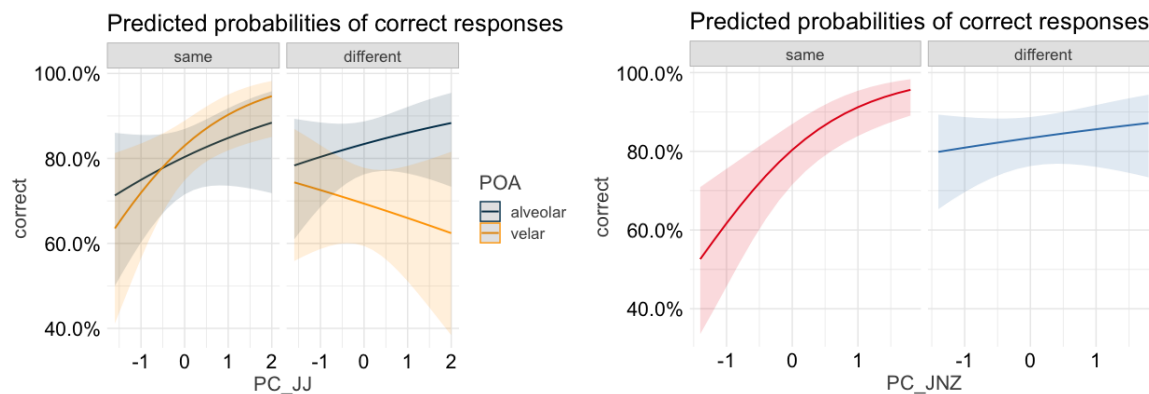


Figure 5.12 Predicted interaction between POA, stimulus pair and PC-JJ (left), and interaction between pair and PC-JNZ (right).

### 5.4.2.3 Native Japanese listeners

Next, discrimination accuracy of native vs. non-native contrasts (i.e., singleton and CVC) by Japanese listeners was examined, using subset data (2,629 tokens). The process to obtain the best fit model is the same as that described above. The best-fit model has two two-way interactions between POA and PAIR, and PAIR and PC-English. Trial number ( $z$ Trial) is as a fixed effect. Along with these factors, SUBJECT and WORD are added as a random intercept in the model. The below is the best model predicting the discrimination accuracy of native vs. non-native contrasts for this group.

```
glmer (correct ~ POA * PAIR + PAIR * PC_English + zTrial + (1+zTrial|subject)
+ (1|word), data = CVC, family = "binomial")
```

The results of the model are presented in Table 5.18. The reference categories are *alveolar same pair*. Once again, the positive effect of trial is detected ( $\beta= 0.31, z=3.40, p<0.001$ ), suggesting listeners' discrimination accuracy increases with each trial. A significant two-way interaction between POA and PAIR is included ( $\beta= -1.43, z=-6.1, p<0.001$ ). This indicates that the discrimination accuracy depends on the relation between place of articulation and type of stimulus pair, and the effects work negatively. This is illustrated in Figure 5.13. The figure on

the left shows the relationship of accuracy and types of stimuli. The x-axes are the four consonant types and the y-axis denotes the percentage of correct trials. The values for velars in the different pairs significantly dropped in comparison to other contexts. This confirms **Prediction 2a (2) and 2b**. When their phonological/phonotactic knowledge leads them to believe that a final consonant is impossible, they show difficulty in discriminating contrasts in the stimuli, specifically in the velar contexts.

The figure on the right shows the probability predicted by the best-fitted model. As shown in these figures, when stimuli of same pairs are presented, there is no difference regardless of place of articulation. On the other hand, when stimuli of different pairs are presented, the accuracy is significantly different between alveolar and velar stimuli. The subset analysis was performed, and it was found that when stimuli of same pairs are presented, the effect of POA ( $\beta= 0.002, z=0.09, p=0.92$ ) is not significant. On the other hand, when stimuli of different pairs are presented, the effect of POA is significant ( $\beta= -1.65, z=-9.1, p<0.001$ ). Thus, the accuracy for the velars is significantly lower than that for the alveolars. Once again, this confirms **Prediction 2b**.

Table 5.18 Model summary for singleton-CVC discrimination by native Japanese listeners

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	1.87068	0.25217	7.418	1.19E-13
POA velar	0.02192	0.21312	0.103	0.918097
pair different	-0.08429	0.16832	-0.501	0.616539
PC_English	0.05732	0.21101	0.272	0.785909
zTrial	0.31536	0.0925	3.409	0.000651
POA velar: pair different	-1.43154	0.2328	-6.149	7.79E-10
pair different: PC_English	0.32106	0.13393	2.397	0.016522

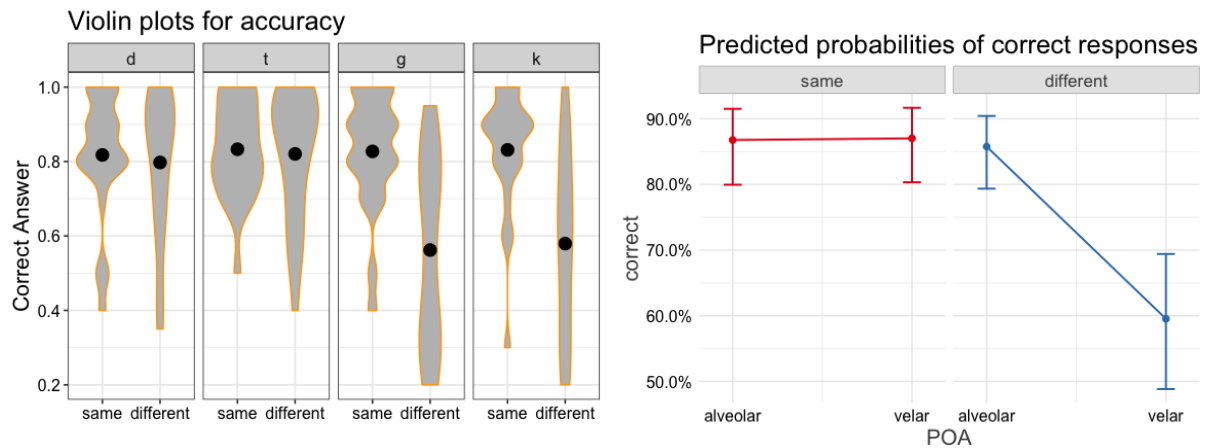


Figure 5.13 Distribution and probability density of discrimination accuracy on singleton–CVC contrast by native Japanese listeners for the raw data (left) predicted interaction between POA and stimulus pair (right).

In addition, a significant two-way interaction between PAIR and PC\_English was detected ( $\beta = 0.32, z = 2.39, p < 0.05$ ). This indicates that the accuracy of discrimination also depends on the relation between type of stimulus pair and PC\_English and the effect of PC\_English is positive in the different pairs. This is illustrated in Figure 5.14, showing the probability predicted by the model. The x-axes present individuals' PC-English scores and the y-axis denotes the percentage of predicted correct trials. It suggests that Japanese listeners with higher exposure to English are more likely to discriminate native/non-native contrasts better than listeners with lower exposure to English, regardless of voicing types or place of articulation. Thus, **Prediction 2c (2)** is confirmed. The performance of Japanese listeners with high exposure to English are better at perceiving between singleton and CVC than that of native Japanese listeners with low exposure to English.

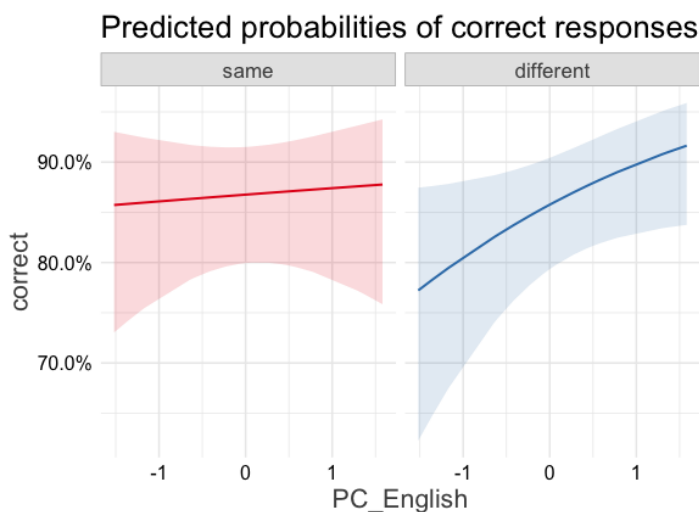


Figure 5.14 Interaction between pair and PC-English.

## 5.5 Discussion

### 5.5.1 Summary of results

#### 5.5.1.1 Singleton-Geminate contrasts

A main purpose of the current study is to determine whether L1 English-L2 Japanese listeners perceive differences between singletons and geminates. It was predicted in **Prediction 1a** that (1) as geminate stimuli are phonemic/phonological categories for native Japanese listeners, they would demonstrate phonemic perception during discrimination of the two stimuli. (2) For L1 English-L2 Japanese listeners, some listeners would be sensitive to distinctions between geminate stimuli although the stimuli are not native-language phonological categories. If listeners could not tap into the phonemic/phonological information, they would show difficulty in discriminating the stimulus contrasts. (3) Monolingual English listeners would not be able to discriminate consonants varying in duration as they do not have phonemic categories according to consonant length. Therefore, it is predicted that performance across language groups would differ.

In addition to this main purpose, mediating effects of phonetic salience on the discrimination of non-native consonant length contrasts were investigated. It was hypothesised in **Prediction 1b** (1) that regardless of groups, if listeners show difficulty in perceiving contrastive pairs, singleton and geminate /k/, /g/ followed by /u/ (i.e., velar stimuli) would be discriminated less well compared to singleton and geminate /t/, /d/ followed by /o/, because the velar context is phonetically less salient than the alveolar contexts. (2) If listeners are quite sensitive to specific acoustic signals, contrasts in voiced stimuli are more likely to be discriminated well than contrasts in voiceless stimuli. In this case, the effect of POA and voicing type would be detected.

As for the effects of the degree of exposure of the individuals to the target language, it was predicted in **Prediction 1c** that the performance of L1 English-L2 Japanese listeners with high exposure to Japanese might be better than that of L1 English-L2 Japanese listeners with low exposure to Japanese.

Regarding **Prediction 1a (1)**, native Japanese listeners demonstrate phonemic perception during discrimination of the two stimuli. Accuracy on the same pairs was above 80%, so therefore nowhere near 100% accurate. This could be due to the nature of the task. The memory decision task forces listeners to hold a number of unfamiliar words until they hear the last word (i.e., the target stimuli) to compare. In addition, the duration of the task was about 30 minutes. Thus, the task demands more working memory resources. If listeners were

distracted and missed the first contrastive pair in the list, it was impossible to compare with a second pair. During the experiment, their accuracy significantly increased (i.e., an effect of z-scored trial), suggesting the listeners strongly orientated their attention to consonant length in stimuli. This indicates that geminates and singletons are stored as the more abstract and distinct category from each other in the mind of Japanese listeners. The listeners discriminate contrast pairs by relying on phonemic/phonological information during the task as predicted. For contrast pairs (i.e., different pairs), neither POA nor voicing types affected their judgment, suggesting no effects of phonetic salience. Interestingly, the results also showed the effects of PC-English, when contrastive pairs were given. Listeners with high exposure to English had more difficulty with perceiving the singleton-geminate contrasts than listeners with low exposure to English. This indicates the influence of L2 when listeners discriminate their own native phonemic categories adversely. The effect of L2 will be discussed more later.

Regarding **Prediction 1a (2)**, for the L1 English-L2 Japanese listeners, some listeners show a sensitivity to phonemic distinctions between geminate stimuli, although the stimuli are not native language phonological categories. During the experiment, accuracy did not increase, suggesting that listeners did not orient their attention to consonant length in the stimuli. That is, listeners were unable to pay attention to particular phonemic features involved in the non-native contrast which enhanced perception. And yet the same time, the mediating effects of phonetic salience were found as **Predicted 1b (1) (2)**. The ability to detect the contrast was significantly decreased by the effects of POA and voicing types, respectively. That is, discrimination of velar stimuli is more difficult than that of alveolar stimuli, regardless of voicing types as predicted. Besides, discrimination of voiceless stimuli is more difficult than that of voiced stimuli, regardless of POA. This provides evidence that phonetic salience of segments may be an influential factor for perceiving consonant length for L1 English-L2 Japanese listeners. This finding is consistent with that of Hardison and Motohashi-Saigo (2010), who found that greater consonant-vowel sonority differences facilitate perception. As expected, voiceless velar stimuli (i.e., [ku] vs. [kku]) was not perceived well as other contexts, indicating that listeners had some difficulty in telling the difference between singleton and geminate in those stimuli. However, listeners were more likely to answer ‘different’ to the different stimuli than the same one. They answered “yes” for 75% of “same” tokens and 50% of “different” tokens in Figure 5.7, showing that “different” tokens are not the same as the “same” tokens. In this case, it seems that listeners can hear the difference at least some of the time.

In regard to **Prediction 1c**, the effects of extra-linguistic factors, PC-JJ (i.e., exposure to Japanese in Japan) and PC-JNZ (i.e., exposure to Japanese in NZ) were detected in a way

which was not expected. While both factors positively enhanced the ability to detect same pairs, the factors did not facilitate ability to detect different pairs. These findings suggest that exposure to the target language is definitely important in the learning and processing of a second language, but length of exposure to the language (Flege, Yeni-Komshian, & Liu, 1999), size of vocabulary (Bundgaard-Nielsen, Best, & Tyler, 2011a, 2011b) or other factors such as individual learners' attitude and motivation toward study the second language (Krashen, 1981) might be related to the development of L2 speech perception and phonological acquisition.

For the present study, a post-hoc test shows that L1 English-L2 Japanese listeners were significantly less accurate on geminate/singleton order than on singleton/geminate order regardless POA or voicing types. This may suggest that some listeners are likely to attend to phonetic detail and have a tendency not to particularly attend the phonological difference. This issue is discussed later in §6.2.3.

Finally, for the monolingual English listeners regarding **Prediction 1a (3)**, when listeners were presented with contrast pairs, performance was significantly more poorly than when listening to stimuli of same pairs regardless of POA or voicing types. However, it is still the case that listeners were more likely to answer 'different' to the different stimuli than the same one, indicating listeners can hear 'something' some of the time. Such difficulty in discriminating the stimulus contrasts was predicted as the listeners do not have phonemic categories according to consonant length. Thus, this specific group of listeners are not able to reliably rely on a language-specific phonemic/phonological information in the way that Japanese listeners do. In addition, listeners were not able to rely on auditory information, presumably due to the rapid decay of acoustic cues during the high memory demand task. In fact, during the experiment, their accuracy did not increase at all, suggesting that listeners were not able to orient their attention to patterns in stimuli. However, the effects of voicing types were detected regardless of velar or alveolar stimuli which was predicted in **Prediction 1b (2)**. This indicates that the phonetic salience of contrasts or mean ratios of geminate to single closure (GC/SC) is effective. Ratio of voiced stimuli were about twice longer than that of voiceless stimuli ([t] 2.45, [d] 5.12, [k] 2.42, [g] 4.22). Therefore, voiced stimuli might be perceived better than voiceless stimuli. Overall, their discrimination accuracy is much lower than that of L1 English-L2 Japanese listeners.

Thus, as predicted, performance across language groups differed. Native Japanese listeners demonstrated phonemic perception in discrimination of the two stimuli. On the other hand, L1 English-L2 learners showed difficulty when discriminating stimulus contrasts dependent on types of stimuli, which was predicted by the mediating effects of phonetic

salience. The positive effect of exposure to Japanese was not found. Monolingual English listeners were not able to discriminate consonants varying in duration as the listeners did not have phonemic categories according to consonant length. However, the effect of mediating of acoustic salience of voicing was found.

### 5.5.1.2 Singleton-CVC contrasts

Another purpose of the study is to determine whether Japanese listeners perceive differences between singletons and CVC /k/, /g/ followed by /u/. As with singleton-geminate contrast, influential factors in the discrimination of non-native consonant length contrasts were investigated. It was predicted in **Prediction 2a** that (1) monolingual English listeners would demonstrate phonemic perception on CVC stimuli that are within their L1 phonological categories. (2) For the Japanese listeners, if their phonology/phonotactic knowledge leads them to believe that a final consonant is impossible, they show difficulty in discriminating the stimulus contrasts. **Prediction 2b** states that regardless of groups, if listeners show difficulty perceiving contrastive pairs, singleton and CVC in the velar context would be discriminated less well compared to in the alveolar context, because the velar context is less phonetically salient than the alveolar contexts.

In addition, the degree of exposure of the individuals to Japanese and/or English might affect their performance. This is based on the assumption that as people learn a second language, it is more likely that they acquire expectations consistent with the structures of the second language. Thus, it was predicted in **Prediction 2c** that (1) performance of L1 English-L2 Japanese listeners with high exposure to Japanese might be lower than that of L1 English-L2 Japanese listeners with low exposure to Japanese. (2) Performance of native Japanese listeners with high exposure to English might be better at perceiving between singleton and CVC than that of native Japanese listeners with low exposure to English.

First, as for **Prediction 2a (1)** for monolingual English listeners, the accuracy on same pairs was above 80%, but were not near 100%. As discussed before, this could be attributed to the high memory demands of the task on the listeners, leading to lower accuracy than in a task with low memory demands. For contrast pairs (i.e., different pairs), the effect of POA was that the accuracy of velar stimuli was significantly lower than that of alveolar stimuli. The phonetic salience of the contrast seems to be an influential factor in perceiving consonant length for even monolingual English listeners regarding **Prediction 2b**. However, they showed very good performance overall on each stimulus type. As predicted, monolingual English listeners

demonstrate phonemic perception on CVC stimuli in their native-language phonological categories. The native phonological knowledge helps them to discriminate the pairs.

L1 English-L2 Japanese listeners showed relatively good performance, similar to the monolingual English speakers. In addition, during the experiment, their accuracy significantly increased (i.e., an effect of z-scored trial), suggesting the listeners orientated their attention to specific patterns in stimuli. However, their discrimination accuracy showed more complex results than monolingual English listeners. Starting with a three-way interaction between POA, PAIR and PC-JJ ( $p < 0.01$ ), a main point of this interaction is that when the velar stimulus pairs were presented, discrimination of the contrast between singleton and CVC was significantly decreased by listeners with high exposure to Japanese in Japan. That is, the listeners exhibited a more Japanese-like discrimination of velar stimuli. On the other hand, exposure helps facilitate listeners to discriminate alveolar contrast pairs. Thus, regarding **Prediction 2b**, phonetic salience of the segment seems to be the influential factor for perceiving singleton-CVC contrasts for the listeners. The negative influence of exposure to second language on native sound contrasts was also found in singleton-geminate contrasts by Japanese listeners, which confirms **Prediction 2c (1)**. The findings might indicate that as more people are exposed to a second language, it is more likely that they are assimilated to the language by the influence of the structures of the second language. The effect of L2 will be discussed in a later section.

Turning now to the effect of exposure to Japanese in NZ, regardless of stimuli types, discrimination accuracy was increased. While exposure to Japanese in Japan influences only specifically discriminating velar stimuli negatively, exposure to Japanese in NZ influences overall discrimination accurately. In fact, these two exposures are essentially different. Exposure to Japanese in Japan involves a certain amount of communication through interactions with Japanese speakers, and generally the language which people normally can hear in Japan is only Japanese. On the hand, exposure to Japanese in NZ in the current study does not involve person-to-person interaction of speaking Japanese. PC-JNZ was related to exposure to Japanese by listeners accessing Japanese websites or watching Japanese TV programs. Thus, the two effects show different results.

Finally, for Japanese listeners regarding **Prediction 2a (2)**, when Japanese listeners were presented with contrast pairs, velar stimuli were perceived less accurately than alveolar stimuli. This indicates that listeners had more difficulty with low sonority stimuli than with high sonority stimuli. The effect of the phonetic salience regarding **Prediction 2b** is consistent across the groups. During the experiment, their accuracy significantly increased, suggesting that the listeners orientated their attention to specific patterns in stimuli. In this group, a



significant interaction of pair with an extra-linguistic factor, PC-English was found. The effect enhanced the discrimination accuracy of Japanese listeners with more exposure to English regardless of voicing types or place of articulation, as anticipated in **Prediction 2c (2)**. Finally, since the accuracy of velar stimuli by Japanese listeners was around 60%, a post-hoc subset analysis in terms of language groups was performed. It revealed a clear effect of the language group in which accuracy of monolingual English listeners was significantly higher than that of Japanese listeners ( $p < 0.05$ ), while between monolingual and learners were not significantly different ( $p = 0.44$ ). Thus, for velar stimuli which are phonetically less salient segments, Japanese listeners have great difficulty discriminating between the stimuli as predicted, however this would be mitigated to some extent by the L2 influence.

In sum, the results of memory decision task were consistent with the prediction that Japanese listeners had difficulty in discriminating singleton-CVC contrasts for the velar stimuli, whereas the discrimination accuracy of the velar stimuli were significantly lower across all groups than that of the alveolar stimuli.

### **5.5.2 Implication for the first experiment**

The aim of the first experiment was to determine whether learners of Japanese acquire sublexicon-specific phonology/phonotactics from Japanese loanword phonology, through the natural language using a statistical learning mechanism. Recall that on the basis of previous studies and corpus data, final stops following lax vowels in English CVC words are borrowed as geminates in loanwords. There are stochastic patterns, such that the likelihood of different voiced geminates varies (i.e., [dd] > [gg] > [bb]). Moreover, voiced geminates are less likely than voiceless geminates in this context. Thus, geminates should have been rated higher than singletons in voiceless contexts (i.e., geminates > singletons). On the other hand, singletons should have been rated higher than geminates in voiced contexts (i.e., singletons > geminates). However, these stochastic patterns were not found in learners' group.

Response patterns in the findings lead to a fundamental question as to whether L1 English-L2 Japanese listeners actually perceived differences between singletons and geminates in the first experiment. Therefore, the study presented in this chapter investigated whether non-native speakers of a language are able to perceive consonant length contrasts that do not occur in their language. In addition, it was examined whether native Japanese listeners have difficulty perceiving differences between singletons and CVC in the same phonological environments, given their surprisingly high levels of acceptance of CVC tokens in the first experiment.

### **5.5.2.1 Implication of singleton and gemination ratings by L2 learners**

The present experiment evaluated the influence of perceptual confusion and provided evidence for the great difficulty that L1 English-L2 Japanese listeners had when discriminating difference between singletons and geminates in the voiceless velar context. However, there is a statistical difference between how they are treating the “same” stimuli and how they are treating the “different” stimuli. That is, listeners were more likely to respond ‘different’ to the different stimuli than the same stimuli. They are able to hear the difference at least some of the time. This clearly reveals that the learners are using some additional knowledge beyond the monolinguals. Meanwhile, the voiced stimuli were discriminated relatively well, especially for the alveolar context. The findings are in line with the results of the first experiment where ratings of singleton and geminates were significantly different in the alveolar contexts than in the velar context.

L1 English-L2 Japanese listeners, then, can distinguish between geminates and singletons to some degree in the present experiment, but did not show differences in their willingness to accept them in the first experiment. Does this pattern of results show that they do not acquire the loanword-specific phonology? In the first experiment (§4.2.2), an overall effect of loanword phonotactics on well-formedness rating was found. Does the present pattern of results contradict that finding? To recall, for the voiced stimuli, learners rely on loanword phonotactics (i.e., a significant positive result), whereas they are less likely to use Japanese overall phonotactics (i.e., a significantly negative impact on ratings). These are reasonable results since the overall phonotactic scores capture only the overall phonotactic variation in the stimuli that are not linked to loanword phonotactics. Voiced geminates appear only in loanwords, and so the ability to discriminate between singleton and geminates is crucial for determining the findings of the first experiment.

The present study show that L2 learners can discriminate voiced singleton and geminates under high demand memory load. In addition to the present results, in the first experiment, learners of Japanese gave low ratings to stimuli with wrong epenthesis, even though the overall Japanese phonotactic score of these stimuli in labial and velar contexts are very high (see §4.1.2.3), which suggests that learners rely on the epenthetic rules. By looking at some aspects of the current results alongside the results from the first experiment, we can conclude the following. The findings support the hypothesis that learners of Japanese have acquired the sublexicon phonology in Japanese to a certain extent through the statistical

learning mechanism. However, fine-grained knowledge of consonant gemination in the target structured loanwords (i.e. [dd] > [gg] > [bb]) was not acquired. Thus, we can conclude that the listeners can discriminate the consonant contrasts to a reasonable degree, but they rated both singleton and geminate stimuli highly in the well-formedness task. This is presumably because, even though learners had knowledge of phonological categories, they had not generalised the target loanword pattern from inputs they encountered.

### **5.5.2.2 Implication of CVC rating by Japanese listeners**

As for native Japanese listeners, they had some difficulty perceiving differences between singletons (i.e., CVCV) and CVC in the velar contexts. We found that the effect of phonetic salience by which in comparison to the alveolar context, Japanese listeners were less likely to discriminate well between [k] and [ku], [g] and [gu], respectively. As we discussed in section §5.2.3, these results indicate that the ‘mishearing’ of CVC as CVCV might be perceptual restoration of a phoneme based on what listeners expect to be there, rather than lack of ability to discriminate two stimuli. That is, listeners’ phonological/phonotactic knowledge (i.e., Japanese linguistic knowledge) leads them to believe that a final consonant is impossible. Thus, listeners rated CVCV (singleton) and CVC forms similarly in the well-formedness task. Same as previous study, only the default epenthesis context (i.e., [u] context), they show difficulty in discriminating contrasts in the stimuli. Hence, the present study strongly suggests that while Japanese syllable structure constraints perception of Japanese listeners, phonetic salience seems to play a role.

In addition, Japanese listeners with higher exposure to English are more likely to discriminate native/non-native contrasts better than listeners with lower exposure to English, which is consistent with the findings in the first experiment. This suggests that influence of L1 linguistic knowledge on their speech perception is mitigated by the increased exposure to the structures of L2.

To conclude, CVC received surprisingly high rates in the well-formedness task, as implicit knowledge of L1 phonology/phonotactics influence listeners’ perception.

## **5.5.3 General implications**

### **5.5.3.1 Effect of acoustic salience**

The study presented in this chapter investigates whether non-native speakers of a language are able to perceive consonant length contrasts that do not occur in their language. As predicted in

**Prediction 1b and 2b**, phonetic similarity plays a mediating role in perceptual confusion. Specific stimuli were difficult to perceive, which might be due to differences in overall acoustic contrasts of CV sequences (e.g., differences between the acoustic properties of /ku/ and /to/). That is, perception of phonological contrasts varies from one to another because certain sequences are phonetically clearer (e.g., Polka, 1991, 1992). The present study showed that overall, when velar stimuli were given, listeners are less likely to accurately discriminate contrastive pairs than when alveolar stimuli were given. Similarly, performance on voiceless stimuli was relatively poorer than on voiced stimuli. This indicates that acoustic signals of the stimuli are the source of perceptual confusion. This is in line with the findings of Hardison and Saigo (2010), who showed that the sonority relationship between the geminate and following vowel plays a role in L2 perception of geminates. The results are also consistent with the argument by Monahan et al. (2009) that Japanese listeners illusorily perceive an epenthetic in environments where [u] is the appropriate epenthetic vowel. The findings suggest that phonetical salience plays a crucial role in perception of consonantal contrasts. The vowel [u] is the shortest vowel and the most susceptible to weakening and deletion among the five Japanese vowels (Hirayama, 2003; Kubozono, 2015; Sagisaka & Tokuhara, 1984 as cited in Irwin, 2011; Shoji & Shoji, 2014), common properties of perceptually weak segments, which is consistent with the view that the epenthetic vowel is the perceptually least salient in the language (Byarushengo, 1976; Fleischhacker, 2001; Kang, 2003; Kenstowicz, 2007; Shinohara, 1997; Steriade, 2001b, 2008). The current experiment showed that the perceptual epenthesis is more likely to occur in environments where [u] is the appropriate epenthetic vowel even in the word-final position.

One might ask whether the findings above are the effect of frequency rather than the effect of acoustic signals, for example, the frequency of exposure to velars is less than that of alveolars. According to Tamaoka and Makioka (2004), frequency of CV sequences between /ku/ and /to/ are not that different on the basis of token frequency as shown in Table 5.19. Tamaoka and Makioka (2004) used a lexical corpus of a total type frequency 341,771 morphemic units and a total frequency of 287,792,797 morphemic units established by Amano and Kondo (2000). The CV /ku/ appears 17,211,261 times, whereas the /to/ has a token frequency value of 17,102,180. For type frequency, /ku/ are counted as 38,359, which is larger than that of /to/ at 27,131. As for voiced sequences, /gu/ and /do/ occur much less frequently in comparison to their voiceless counterparts. If the frequency is the influential factor on perceptual confusion, discrimination of voiced stimuli should have been poorer than that voiceless. Thus, the effects of acoustic signals would better account for perceptual confusion

than the effects of frequency. The findings suggest perceived similarity and differences stemmed from the phonetic level.

Table 5.19 Frequency counts of morae based on Tamaoka and Makioka (2004)

	CV	Token frequency	Type frequency
voiceless	/ku/	17,211,261	38,359
	/to/	17,102,180	27,131
voiced	/gu/	1,210,586	8,951
	/do/	5,306,817	12,158

The findings of the previous study (the first experiment) and this present study taken together, will be discussed in regard to theory throughout the next chapter.

### 5.5.3.2 Effects of L2 language experiences on L1 perception

While not directly related to the primary question that we set out to address, the experiment showed that an extra-linguistic factor – exposure to second language – might influence how native speakers of a language perceive their native sound contrasts. L2 experiences adversely affected the native sound perception for both groups, when velar stimuli were given to them. The trend was that listeners with more exposure to their second language exhibited L2-like discrimination of native phonemic contrast pairs. That is, Japanese listeners with more exposure to English exhibited more L1 English-L2 Japanese like discrimination of singleton-geminate contrasts. Similarly, L1 English-L2 Japanese listeners with more exposure to Japanese in Japan exhibited Japanese-like discrimination of singleton-CVC velar stimuli. As for the L1 English-L2 Japanese listeners, exposure to Japanese in Japan (i.e., PC-JJ) is not related to their Japanese proficiency at all. PC-JJ scores were extracted from questions related to visiting Japan. On the other hand, PC-English scores for Japanese listeners were extracted from the questions about self-reported levels of proficiency in English and the duration of their time in NZ. Thus, questions were not identical for the two groups and the differential exposure in their language experiences. For that reason, the effects on discrimination of contrastive sounds cannot be compared directly.

However, both groups exhibited similar patterns when perceiving their native sound contrasts, indicating a cross-language effect in phonological and phonetic levels. Previous studies on language processing in bilinguals and L2 learners found that speakers' L2 influence to access their L1, including lexical access (e.g, Bice & Kroll, 2015; Ivanova & Costa, 2008;

Levy, McVeigh, Marful, & Anderson, 2007; Marian & Spivey, 2003b), phonetic production (e.g., Chang, 2012, 2013; Flege, Schirru, & MacKay, 2003; Linck et al., 2009; Major, 1992; Mora & Nadeu, 2012) and speech perception (e.g., Mora & Nadeu, 2012; Tice & Woodley, 2012). For example, a five-week longitudinal study by Chang (2012) showed that English stop consonants and vowels production were phonetically assimilated to those of Korean in early L2 acquisition of L1 English-L2 Korean speakers. Since speech production is strongly related to incoming auditory information of L2, it is reasonable to expect that L2 experience influences the way learners perceive their own native sounds. Mora and Nadeu (2012) examined the effect of L2 (Spanish) on the phonetic perception of a Catalan mid-vowel contrast by two groups of Catalan-Spanish bilinguals, according to the degree of their daily exposure/use of Catalan. The results showed frequent Spanish users discriminated the Catalan contrast less accurately and more slowly than the other group, suggesting the amount of exposure/use has an effect on the backward influence of L2 on L1 perception. Thus, the present findings are in line with the effect of L2 on the categorical perception.

Although there is an abundance of literature on how native language experience influences adult learners' speech perception of non-native sound contrast (e.g., Dupoux et al., 1999; Dupoux et al., 2011; Ingram, 1997; Kabak & Idsardi, 2007; MacKain, Best, & Strange, 1981; Sheldon & Strange, 1982; Takagi & Mann, 1995; Yamada & Tohkura, 1992), and the phenomena has been discussed theoretically (e.g., Best, 1994; Best, 1995; Best & Tyler, 2007; Flege, 1999; Kuhl, 1991), until recently the influence of second language on the perception of native language has been relatively less discussed. Therefore, the current findings might shed light on the effects of L2 experience on the categorical perception of L1. Since for both groups, their discrimination accuracy significantly increased during the experiments, they oriented their attention to the target phonological features in the stimuli. Therefore, it is speculated that as more people are exposed to a second language, it is more likely that they are assimilated to the language by the influence of the structures of the second language.

### **5.5.3.3 Attention control**

Second-language experience also enhanced the ability to discriminate non-native contrasts for Japanese listeners, but not for L1 English-L2 Japanese listeners. For the Japanese listeners, listeners with more exposure to English exhibited higher discrimination accuracy than listeners with lower exposure to English. However, L2 learners did not show similar patterns. This discrepancy may be due to 'attention control' which is the cognitive ability to shift efficient

attention between task-relevant and task-irrelevant information (Isaacs & Trofimovich, 2011; Rosen & Engle, 1998). During the experiment task, the discrimination accuracy of non-native contrasts only by Japanese listeners significantly increased, suggesting they controlled their attention to task-relevant information. As mentioned before, attention control plays an important role in the discrimination of sounds in listeners' perception to detect a specific phonetic cue in stimuli, and attention enhances learning (e.g., Asano, 2018; Chen, Best, & Antoniou, 2019; Darcy, Mora, & Daidone, 2014; Guion & Pederson, 2007; McCandliss, Fiez, Protopapas, Conway, & McClelland, 2002). In addition, the effect of attention seems to play a role for generalising linguistic structure during statistical learning tasks (Toro, Sinnett, & Soto-Faraco, 2005, 2011). In the present study, because the listeners detected important features in stimuli, phonological knowledge of a second language might assist the perceptual abilities of the listeners to discriminate pairs of stimuli. That is, listeners could tap into the phonemic/phonological information of the second language. In other words, if L1 English-L2 learners were able to pay attention to the consonants of varying durations in the stimuli, their knowledge of a second language might have enhanced their discrimination accuracy.

#### **5.5.3.4 The importance of type of task**

In terms of the type of task, the results of this study taken together support the view of Werker and Logan (1985) that categorical perception is dependent on experimental tasks or conditions. When a task condition demands high memory load, phonetic or phonemic processing are required rather than auditory processing. The present task condition is more similar to everyday oral communication from Werker and Logan's perspective, increasing listeners' memory load. Therefore, it would predict that participants in the present experiment should have more difficulty to discriminate consonant length contrasts than the experiments outlined in the literature reviews, in which naïve listeners were capable of discriminating the contrasts accurately enough. Indeed, in the present experiment, monolingual English listeners' discrimination accuracy did not reach 50%, particularly showing difficulty in voiceless contexts. This is because, firstly, listeners could not access the auditory temporal cues differentiating contrasts. Listeners' attention was intentionally manipulated by a more complex task, and the listeners were forced to engage their memory for more than 1500ms which entails phonemic processing for listeners. Auditory memory decayed, while listeners needed to hold in memory a number of unfamiliar words until they heard the last word (i.e., the target stimuli) to compare. That is, listeners could not perform the auditory level of processing. This finding is consistent with the view that auditory perception is possible only in short ISI conditions

(Crowder, 1982; Pisoni, 1973; Werker & Logan, 1985). Secondly, participants in the current study were not given explicit instruction about the nature of contrasts. In the previous studies (Asano, 2018; Hisagi & Strange, 2011; Porretta & Tucker, 2015), discrimination of non-native contrasts was enhanced by giving information about what to listen for. This made naïve participants focus directly on the durational contrasts of singleton and geminate consonants, which enhanced the phonetic level of processing rather than auditory processing. The present results indicate that without paying attention to the target segments, phonetic mode processing is difficult for naïve listeners at least under high demand memory load.

Additionally, the present results indicate that even under increased task demands, native speakers demonstrate the level of phonemic process according to phonological categories in their native language without explicit information about particular cues to discriminate contrasts. In light of speech processing strategies, these findings are consistent with categorical perception that are dependent on the task conditions and memory demands. Thus, the present study revealed the importance of the type of experimental task used to investigate how accurately listeners perceive non-native contrasts in relation to the processing factors.

## **5.6 Conclusion**

In conclusion, the findings provide supporting evidence that L2 learners can acquire sublexicon phonology and phonotactics in Japanese to a certain extent without being explicitly taught. This is supported by discriminating of two distinct stimuli under a task with high memory load and the findings from the first experiment. Additionally, comparison to English-speaking monolinguals the L2 learners are apparently using some additional knowledge beyond the monolinguals. Most importantly, taken together, the findings of two experiments support that such acquisition is possible through the statistical learning mechanism. In addition, implicit knowledge of L1 phonology/phonotactics seems to influence Japanese listeners' perception on CVC which received surprisingly high rates in the well-formedness task. This study also shows regardless of groups, listeners exhibited difficulty in perceiving contrastive pairs, singleton and geminate /k/, /g/ followed by /u/ (i.e., velar stimuli) compared to singleton and geminate /t/, /d/ followed by /o/ because the velar context is acoustically less salient than the alveolar context. In addition, L2 experiences adversely affected their native sound perception for both groups, when velar stimuli were presented to them.



There is a remaining question: Why do native Japanese speakers successfully learn gradient rules, but L2 learners do not? This question is discussed in the following section after summarising the predictions and findings in two experiments.

# Chapter 6

## Discussion and Conclusion

### **6.1 Summary of Research Question, Predictions and Findings**

This thesis focuses on the L2 language acquisition of Japanese sublexicon phonology in relation to statistical language learning. Two experiments were conducted to investigate whether L2 learners can learn the loanword phonotactics/phonology of Japanese through the experience of using and/or passive exposure to Japanese lexical stratification. Using two loanword phonological regularities as a case study, the first experiment was designed to explore listeners' phonotactic/phonological knowledge of nativised loanwords in Japanese using a well-formedness task. The second experiment was designed to support the findings in the first experiment which tested whether listeners were able to discriminate non-native consonantal contrasts.

In this chapter, I will briefly summarise each experiment and crucial findings by reviewing the research questions and predictions in §6.1, moving on the discussion which addresses a remaining question and some implications of the present study in §6.2, and pointing out the limitations of the current study. Finally, I will draw conclusions in §6.3.

#### **6.1.1 Summary of confidence-rating task (well-formedness task)**

The confidence-rating task in Chapter 4 investigated the extent to which native and non-native speakers implicitly acquire loanword phonology/phonotactics in a natural language, using the statistical learning mechanism. The study focused on two loanword phonological regularities: epenthetic vowels and consonant gemination. For epenthetic vowels, there are categorical constraints dictating which vowels should be used. For geminates, within the loanwords, there are stochastic patterns, such that the likelihood of different voiced geminates varies. In addition, voiced geminates are less likely than voiceless geminates in the loanwords. I assumed that forming general rules and detecting stochastic patterns are both employed via the same mechanism. The question of interest was whether language users can extract the patterns of epenthetic vowels from instances of distributional contexts of loanwords in the natural language, generalizing patterns to novel instances without supervision. The other question was whether learners are sensitive to the fine-grained patterns of voiced geminates that only occur in the loanwords.

Thus, in Chapter 4, a fully-crossed online investigation into adaptation of English final consonants was conducted with 22 English-speaking learners of Japanese, 20 English-monolingual and 20 Japanese listeners. There were 60 CVC-structured English words, each of them having four different pronunciations: CVC (*pip*), singleton (CVCV; *pipu*), geminate (CVCCV; *pippu*) and wrong epenthesis (*pippo*). The stimulus word was presented orthographically (e.g., <pip>) to participants as they simultaneously heard one of the pronunciations. Participants judged whether the pronunciation they heard was how the word would be pronounced if this was a Japanese word, rating how confident they are on a scale of 1-5.

This experiment demonstrates that both native speakers of Japanese and L2 learners have knowledge of epenthetic rules. Native speakers of Japanese show greater sensitivity to the fine-grained patterns of voiced geminates, but L2 learners of Japanese did not show similar levels of sensitivity. For the L2 learners of Japanese, learners with more exposure to Japanese in NZ show their sensitivity towards voiceless geminations but their response patterns are not consistent with stochastic distributional patterns in loanword gemination. English-monolinguals with more exposure to Japanese were more likely to rate geminates higher. The findings with research questions and predictions presented in Chapter 2 appear below.

**RQ 1:** Is it possible that a sublexicon phonology of a language is learned from exposure to the target language?

**Prediction 1:** It is predicted that the learning of phonological rules is possible without being taught. If this is possible, acquisition of phonological rules would differ depending on the degree to which a learner is exposed to Japanese. This is based on the assumption that participants who have more exposure to Japanese have a reasonable level of phonological knowledge due to their accumulated Japanese lexicon.

Prediction 1 was partially supported by the findings in the experiment. As predicted, the learning of phonological rules was possible without being taught. However, acquisition of phonological rules did not differ depending on the degree to which a learner was exposed to Japanese. This is presumably related to the size of vocabularies. Even learners who have 4000 words in their Japanese vocabulary are likely to know only around 300 loanwords (see §2.3.3.2). We will discuss this speculation in §6.2.4.

**RQ2:** If any, what rules are implicitly learned?

**Prediction 2:** For epenthetic vowels (Rule A), there are categorical constraints dictating which vowels should be used. For geminates within our target sublexicon (i.e. loanwords), there are stochastic patterns, such that the likelihood of different voiced geminates varies. Moreover, voiced geminates are less likely than voiceless geminates in this context. Thus, it is predicted that L2 learners of Japanese are more likely to acquire Rule A (i.e., epenthetic vowel) than Rule B (stochastic patterns of consonant gemination) by exploiting their lexicon. This is based on an assumption that categorical rules are more readily learned than gradient ones. However, the epenthetic vowel [u] occurs more frequently in comparison to the other vowel [o], which occurs only after alveolar stops /t, d/. Thus, it would be possible that language users overgeneralise an epenthetic rule in which [u] can be used in any context.

As predicted, L2 learners of Japanese have acquired epenthetic vowels (i.e., Rule A) rather than stochastic patterns of consonant gemination (Rule B). Thus, categorical rules have been more readily learned than gradient ones. This was expected from previous works showing that adult learners produce categorical phonological sounds more accurately than gradient ones (Shea & Curtin, 2011). In addition, in order to detect gradient patterns, generally more data is required which entails greater lexical knowledge. The well-formedness experiment demonstrated that language users overgeneralise an epenthetic rule in which [u] can be used in any context.

**RQ 3:** Are L2 learners with high exposure to Japanese able to acquire fine-grained knowledge regarding the effects of voicing and place of articulation (POA)?

**Prediction 3:** L2 learners with high exposure to Japanese would be expected to be able to acquire fine-grained knowledge regarding the effects of voicing. The acquisition of this gradient pattern depends on the degree of exposure individuals experience, and the degree of statistical support of the rule. For the effect of POA, [d] is most likely to geminate, while [g] and [b] will have lower ratings for gemination.

The results revealed that L2 learners with high exposure to Japanese have not acquired fine-grained knowledge regarding the effects of voicing and place of articulation (POA). On the

other hand, native Japanese speakers showed their fine-grained knowledge regarding gemination in loanwords, suggesting the size of vocabularies might play a significant role in statistical language learning, detecting and tracking likelihood of gemination.

**RQ 4:** Does the language's overall statistical patterns influence learners' response patterns? Specifically, would participants be biased in responding with the most expected pattern in the language rather than with observed patterns in the Japanese loanwords?

**Prediction 4:** Participants who do not have access to the knowledge of loanword phonology may access the statistical patterns in the lexicon for their responses, in order to find the most frequent phonotactic patterns in the language. In that case, singleton would be selected rather than geminates as the overall frequency of geminates is lower in all contexts.

The first experiment demonstrated that learners rely on loanword phonotactics during the experiment rather than Japanese overall phonotactics for both voiced and voiceless stimuli. Especially, for the voiced stimuli, learners rely on loanword phonotactics (i.e., a significant positive result), whereas they are less likely to use Japanese overall phonotactics (i.e., a significantly negative impact on ratings).

Thus, we found that learners of Japanese have acquired the sublexicon phonology in Japanese to a certain extent through the statistical learning. There are remaining questions: Why did singletons (i.e., CVCV) and geminates (i.e., CVCCV) in labial and velar contexts receive equivalently high-ratings from L1 English-L2 Japanese listeners? Despite the fact that distributional pattern of geminates in the target structure (i.e., English CVC words) differ from that singleton in loanwords. Why did native Japanese speakers give their surprisingly high levels of acceptance of CVC tokens in this experiment? In order to address these important questions, the second perception experiment was conducted. The second experiment is summarised in the next section.

### **6.1.2 Summary of auditory memory decision task**

Chapter 5 presented another cross-linguistic experiment, an auditory memory decision task which builds on the results of the first experiment to explore the role of phonotactics on perception. This study aimed to examine the degree to which non-native speakers can perceive contrasts that do not occur in their native language – singleton/geminate contrasts for English speakers, and CV/CVC contrasts for Japanese speakers. It also investigates the degree to which success in this task is mediated by phonetic salience of the particular contrast, and by the individual's previous language experience. It was hypothesised that regardless of groups, if listeners show difficulty in perceiving contrastive pairs, singleton and geminate /k/, /g/ followed by /u/ (i.e., velar stimuli) would be discriminated less well compared to singleton and geminate /t/, /d/ followed by /o/, because the velar context is acoustically less salient than the alveolar contexts.

Twenty-two L1 Japanese-L2 English, twenty-three L1 English-L2 Japanese and seventeen monolingual English listeners listened to a 240 five-word audio list (Appendix F), containing a stimulus pair of either same or different pairs in random order. In each trial, participants heard a list of four words followed by a 300ms beep. After the beep, another word would be heard. They were asked to judge whether the word after the beep was in the list words before the beep by clicking on one of two options, 'Yes' or 'No' on the computer screen. As in the first experiment, levels of exposure to Japanese/English were measured using PCA.

The findings with research questions and predictions for thesis overarching objective presented Chapter 5 is below.

#### **Prediction 1a: Discrimination of singleton – geminate contrasts**

- (1) As for singleton-geminate contrasts, geminate stimuli are phonemic/phonological categories for native Japanese listeners, so that they would demonstrate phonemic perception of discrimination between the two stimuli.
- (2) Considering L1 English-L2 Japanese listeners, some listeners would show a sensitivity to phonemic distinctions between geminate stimuli, although the stimuli are not-native language phonological categories. If they could not tap into the phonemic/phonological information, they would show difficulty in discriminating between the stimulus contrasts.

- (3) Monolingual English listeners would not be able to discriminate between consonants varying in duration as the listeners do not have phonemic categories according to consonant length.

Therefore, it is predicted that performance across language groups would differ.

Prediction 1a was supported by the findings in the experiment. Firstly, as predicted, performance across language groups differed. (1) Native Japanese listeners demonstrated phonemic perception of discrimination between the two stimuli. On the other hand, (2) for L1 English-L2 Japanese listeners, the results show that they are able to differentiate between singletons and geminates depending on phonological contexts; voiced alveolar stimuli (i.e., [do] vs. [ddo]), voiced velar (i.e., [gu] vs. [ggu]) and voiceless alveolar stimuli (i.e., [to] vs. [tto]) were above 50% and voiceless velar stimuli (i.e., [ku] vs. [kku]) was heard as different below 50% of time (3) Monolingual English listeners showed great difficulty in discriminating between consonants varying in duration. When comparing learners' results with that of monolingual ones, learners' discrimination ability is apparent. As predicted, performance across groups differed.

### **Prediction 1b: Mediating effects of phonetic salience on perception of geminates**

If specific stimuli were difficult to perceive, this might be due to differences in overall acoustic contrasts of CV sequences (e.g., differences between the acoustic properties of /ku/ and /to/). That is, perception of phonological contrasts varies from one to another because certain sequences are phonetically clearer (e.g., Polka, 1991, 1992).

- (1) Regardless of groups, if listeners show difficulty perceiving contrastive pairs, singleton and geminate /k/, /g/ followed by /u/ (i.e., velar stimuli) would be discriminated less well compared to singleton and geminate /t/, /d/ followed by /o/, because the velar context is less phonetically salient than the alveolar contexts.
- (2) If listeners are quite sensitive to acoustic signals, contrasts in voiced stimuli are more likely to be discriminated better than contrasts in voiceless stimuli. This is based on the degree of acoustic difference in the voiced and voiceless stimuli; mean ratios of geminate to single closure (GC/SC) are about twice for voiced stimuli.

The ability to detect the contrast by L2 learners was significantly decreased by the effects of POA and voicing types, respectively. That is, discrimination of velar stimuli is more

difficult than that of alveolar stimuli, regardless of voicing types as predicted. Besides, discrimination of voiceless stimuli is more difficult than that of voiced stimuli, regardless of POA. As well as L2 learners, the effects of voicing types were detected for English monolingual listeners. This provides evidence that acoustic salience of segments may be an influential factor for perceiving consonant length for L1 English-L2 Japanese listeners.

**Prediction 1c: The effects of the degree of exposure of the individuals to the target language on geminate perception**

Greater exposure to a target language presumably leads to an increased ability to identify the relevant phonological contrasts. Thus, levels of exposure to Japanese/English might affect their performance. Hence, it would be predicted that the performance of L1 English-L2 Japanese listeners with high exposure to Japanese might be better than that of L1 English-L2 Japanese listeners with low exposure to Japanese.

For Japanese listeners, the effects of PC-English (i.e., exposure to English in NZ) was found, when contrastive pairs were given. Listeners with high exposure to English had more difficulty with perceiving the singleton-geminate contrasts than listeners with low exposure to English. This indicates the influence of L2 experiences when listeners discriminate their own native phonemic categories adversely.

For L2 learners, PC-JJ (i.e., exposure to Japanese in Japan) and PC-JNZ (i.e., exposure to Japanese in NZ) were detected in a way which was not expected. While both factors positively enhanced the ability to detect same pairs, the factors did not facilitate ability to detect different pairs. These findings suggest that exposure to the target language is definitely important in the learning and processing of a second language, but other factors might be related to the development of L2 speech perception and phonological acquisition (see Discussion in §5.5.1.1).

**Prediction 2a: Discrimination of singleton – CVC contrasts**

- (1) As for singleton-CVC contrasts, monolingual English listeners will demonstrate phonemic perception of CVC stimuli that are within their L1 phonological categories.
- (2) As for Japanese listeners, if their phonological/phonotactic knowledge leads them to believe that a final consonant is impossible, they would show difficulty in discriminating contrasts in the stimuli. Therefore, it is predicted that performance across language groups would differ.



As predicted, (1) monolingual English listeners demonstrated phonemic perception of CVC stimuli within their native-language phonological categories. (2) As for Japanese listeners, they show difficulty in discriminating CVC and singleton in the velar context in comparison to the alveolar contexts.

**Prediction 2b: Mediating effects of phonetic salience on perception of CVC**

Regardless of groups, if listeners show difficulty perceiving contrastive pairs, singleton and CVC in the velar context would be discriminated less well compared to in the alveolar context, because the velar context is less phonetically salient than the alveolar contexts.

The effect of phonetic salience was found across three groups. For contrast pairs (i.e., different pairs), the effect of POA was that the accuracy of velar stimuli was significantly lower than that of alveolar stimuli. The phonetic salience of the contrast seems to be an influential factor in perceiving consonant length for even monolingual English listeners. Hence, the present study suggests that while Japanese syllable structure constraints perception of Japanese listeners, phonetic salience seems to play a role.

**Prediction 2c: The effects of the degree of exposure of the individuals to the target language on CVC perception**

(1) It would be predicted that the performance of L1 English-L2 Japanese listeners with high exposure to Japanese might be lower than that of L1 English-L2 Japanese listeners with low exposure to Japanese. This is based on the assumption that as more people are exposed a second language, it is more likely that they acquire expectations consistent with the structures of the second language. That is why in general, advanced L2 learners perform well in perceptual discrimination tasks.

(2) Similarly, it would be also predicted that the performance of Japanese listeners with high exposure to English might be better at perceiving between singleton and CVC than that of native Japanese listeners with low exposure to English.

(1) When stimuli of different pairs are presented, native contrasts for velar are rather difficult for listeners with high PC-JJ (i.e., exposure to Japanese in Japan) to discriminate, in comparison to that for alveolar. This suggests that exposure to Japanese in Japan influences discrimination of native sound contrasts negatively. On the other hand, the effect of PC-JNZ (i.e., exposure to Japanese in NZ) does not show an interaction with place of articulation and the positive effects

on both the same and different pairs. (2) Japanese listeners with higher exposure to English are more likely to discriminate native/non-native contrasts better than listeners with lower exposure to English, regardless of voicing types or place of articulation.

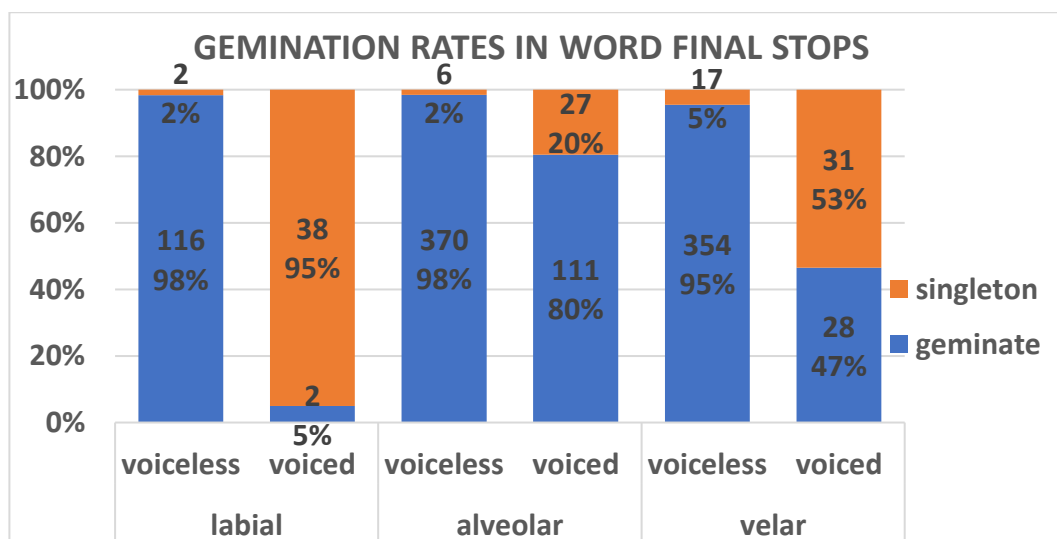
The present work supplies conclusive information regarding the discrimination ability of singleton/geminate contrasts. L1 English-L2 Japanese listeners can distinguish between geminates and singletons to some degree in the present experiment, but did not show differences in their willingness to accept them in the first experiment. This is presumably even though learners had knowledge of phonological categories, they had not generalised the target loanword pattern from inputs they encountered.

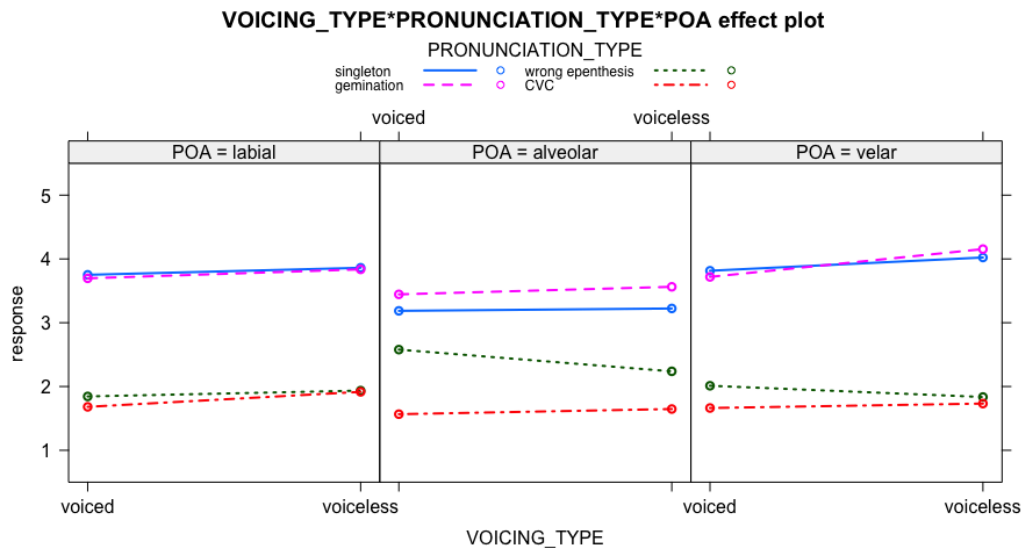
CVC received surprisingly high rates by Japanese listeners in the first experiment, because L1 phonology and acoustic salience leads them perceptual phonemic restoration in the labial and velar context in the previous experiment.

### 6.1.3 Bringing corpus study, well-formedness judgement and perceptual discrimination together

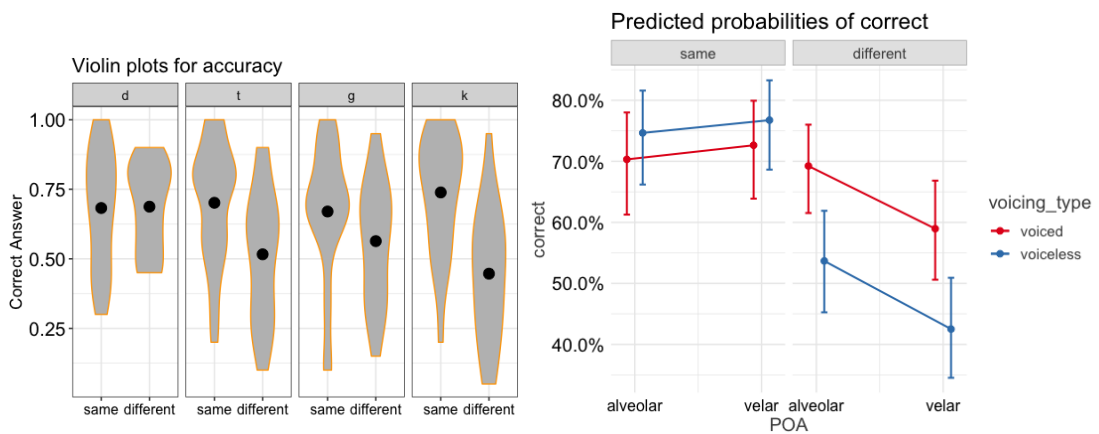
The goal of this section is to discuss the differences in learners' judgments between places of articulation in the first experiment and the second experiment and how they relate to each other, specifically in connection to the relationship between alveolars and velars. Therefore, the results of the three parts of the thesis are considered as a whole. For ease of reference, Figure 2.2, Figure 4.16 and Figure 5.7 from previous chapters are repeated here as Figure 6.1 (a), (b), and (c) respectively.

(a)





(b)



(c)

Figure 6.1 Results from three different analysis regarding POA. (a) Top panel: Frequency of geminates and singletons for source words with a word-final stop following a lax vowel in the BCCWJ. (b) Middle panel: Plots of the three-way interaction between VOICING TYPE, PRONUNCIATION TYPE, and POA in the English-speaking learners of Japanese dataset. (c) Bottom panel: Distribution and probability density of discrimination accuracy on singleton-geminate contrast by learners of Japanese (left) and predicted interaction between POA, voicing type and stimulus pair (right). (Note that Figure 2.2, Figure 4.16 and Figure 5.7 are repeated here for ease of reference)

Figure 6.1 (a) shows the frequency of the singleton/geminate stop occurrence in loanwords whose source words have stop consonants in word-final position after lax vowel based on the BCCWJ data. In the corpus study, when the word-final consonants in the source words are voiceless stops, geminates are preferred as the frequencies of the occurrence of the geminates are almost 100% in nativised loanwords. On the other hand, in the case of voiced stops, occurrences of geminates depend on the place of articulation. Thus, it was expected that the well-formedness ratings in the first experiment would reflect the probability of gemination

in loanwords, if learners implicitly acquire gemination rules in loanwords. (a) voiceless geminates would be preferred over voiced geminates, and (b) voiced geminates would be ranked in the order [dd] > [gg] > [bb]. However, for this prediction to completely follow, we also need to know whether listeners can actually distinguish between the geminate and non-geminate forms. This was the basis of the second experiment (exp 2).

### **Are Voiceless Geminates preferred over Voiced Geminates?**

With respect to the voicing type, the predicted pattern is not observed in the well-formedness rating task. The model predictions in Figure 6.1(b) show that ratings do not substantially change with voicing type. That is, for example, when a voiceless stimulus ‘*peck*’ was given, a geminate /pekku/ is not rated particularly more well-formed than a singleton /peku/. Similarly, when a voiced stimulus ‘*keg*’ was given, a singleton /kegu/ is not rated particularly well-formed than /keggu/.

Despite the fact that voiceless geminates are much more frequent than voiced geminates (see panel a), geminates in the voiceless context are not rated particularly more well-formed than geminates in the voiced context in the first experiment (panel b). When we look at the second experiment (panel c), we see that people can hear the voiced geminates a bit more easily than the voiceless geminates. If anything, this might lower the voiced geminates’ ratings even more, because they would sometimes be in their own separate category, whereas the voiceless ones would more often be heard as a bigger, collapsed category also containing singletons. This therefore seems to be a contradiction between what learners can hear and the statistical patterns they have learned. This suggests that learners have not separately generalised gemination rules according to voicing type. Whatever they have learned about the statistical occurrence of geminates, it is not granular enough to have incorporated the different patterns across voiced and voiceless forms.

### **Are voiced geminates ranked in the order [dd] > [gg] > [bb]?**

With respect to place of articulation effects within voiced forms, the frequency effects led us to predict the geminate preferences would be ordered [dd] > [gg] > [bb]. In terms of discrimination (exp 2), we found that alveolar geminates are actually more perceptually distinguished from singletons than velar geminates are (Figure 6.1 (c)). Both of these would then lead us to expect high ratings for [dd] in the well-formedness rating task than [gg]. In the task (Figure 6.1 (b)), geminates in the voiced alveolar context (e.g., /meddo/, /hiddo/) are rated as more well-formed than singletons (e.g., /medo/, /hido/), which is consistent with gemination

frequency in the panel (a). This is consistent with the idea, from the second experiment, that the alveolar geminates can be reasonably well heard. This differs from the velar consonants, where the well-formedness ratings do not distinguish geminates and singletons, consistent with the interpretation, supported by the second experiment that velar geminates are less well heard. Thus, in this sense, the results of three studies do not create any contradictions.

However, both singleton and geminate alveolars are actually judged less well-formed in comparison to that of velars in the first experiment. This is counter to the predicted direction, since alveolars are more frequent. In addition, while gemination is more preferred in the alveolar context (relative to singletons), both singleton and geminate forms are preferred to the same degree in a velar context, and both receive higher ratings than the alveolars. This is despite the fact that alveolars are more frequent than velars in this context. Perhaps geminates and singletons are equally preferred in a velar context because listeners cannot distinguish between them very well?

The results of the discrimination task for learners in Figure 6.1(c), suggest that this might be the case. Listeners have much more difficulty distinguishing between geminates and singletons in the velar context than the alveolar context. The reason velar singletons and geminates may therefore be rated more highly than the alveolar ones, may be that because for some learners they are still the same category. Hence, a more frequent category is rated more highly, and tokens of both geminates and singletons are in the same (large) category for those learners. In this case, the category is very likely to be cognitively associated with singletons. This is because singletons are learners' native category and geminates are not.

The default epenthetic vowel [u] might also play a contributing role. In this case, some learners may think of the default vowel as a more appropriate vowel than the contextual epenthetic vowel [o], as discussed in §4.5.2. However, ratings for wrong epenthesis are significantly lower than that of singleton and gemination in Figure 6.1 (b). Thus, many learners tend to implicitly acquire the contextual epenthetic vowel. Perhaps it is acquirable because it is less complex than the geminate/voicing interaction required to learn the geminate rule. In addition, it was not only learners but also native Japanese speakers who rated gemination lower in the voiceless alveolar context in comparison to that of velar, even though the contextual epenthetic vowel [o] was almost invariably used in nativised loanwords. To the best of my knowledge, only one instance in which the [u] epenthetic vowel in the alveolar context was found in loanwords as /insuturumento/ 'instrument'. Despite the fact, participants who rated alveolar voiceless geminates lower than that of velar might be avoiding sounding "too nativised" (e.g., Davidson, 2010). That is, participants who do not have the generalisation of epenthetic

rules, they may seek the best pronunciation to comply with Japanese. In the first experiment, it was clearly mentioned that the research is interested in how foreign words are pronounced, when they are borrowed in Japanese. Thus, those participants who do not generalise the epenthetic rules may compromise between well-formedness [to, tto, do, ddo] and more foreignness like sounds [tu, ttu, du, ddu].

## **Summary**

When considering findings from these three results together, we can see that learners' ability to discriminate between singletons and geminates in natural settings may influence statistical learning. This interpretation will be discussed in §6.2.3.

The next section provides general discussion, implications and research limitations. We return to the literature reviewed in the Introduction along with other relevant literature, and discuss why L2 learners are not able to generalise the gradient geminate patterns in loanwords, while native speakers of Japanese successfully learn gradient rules.

## **6.2 General discussion**

### **6.2.1 Influence of L1 on L2 statistical learning: Lexical activation and priming**

As mentioned in §1.3, statistical learning and rule learning are based on the same single domain-general mechanism (Aslin, 2017; Aslin & Newport, 2012, 2014). Studies of artificial language learning showed that infants and adults are able to acquire new languages after short exposure (Aslin et al., 1998; K. E. Chambers et al., 2003; Maye et al., 2002; Onishi et al., 2002; Saffran, Aslin, et al., 1996; Saffran, Newport, et al., 1996). That is, statistical language learning is not only a property of native speakers, but L2 and L3 also tap into the same general cognitive mechanism (Mitchell, Myles, & Marsden, 2019). Importantly, Morita (2018) proved the learnability of sublexicons from naturalistic data in cases where different subphonological systems coexist within the same language. His study was grounded in a Bayesian learning-based computational clustering model that was applied to Japanese and English words from corpora, predicting etymological lexical subclass from segmental phonotactics. Most importantly, his study suggests that language users should be able to learn sublexicon-specific phonotactics based on the same kind of statistical probabilities that computers analyse from language users' accumulated lexicons.

The current study demonstrated that phonotactics/phonology of sublexicons was learned to some extent by L1 English-L2 Japanese speakers without instruction, when learning in natural language settings. L2 learners in the current study were native speakers of English and the target language for loanwords is their native language. Therefore, they should have an advantage in perceiving the English input which inevitably overlaps with the phonological form of the host language (i.e., English loanwords in Japanese). Since most modern loanwords in Japanese are of English origin, distributional patterns of the quality of epenthetic vowels in loanwords might be primed, which might have helped L2 learners' performance.

To account for this possibility, we first consider studies of spoken-word processing that show that listeners' L1 lexicons are activating during listening to L2 languages (e.g., Marian, Blumenfeld, & Boukrina, 2008; Marian & Spivey, 2003a, 2003b; Schulpen, Dijkstra, Schriefers, & Hasper, 2003; Weber & Cutler, 2004). For example, for L1 Russian-L2 English speakers, while hearing an English word 'shark', the Russian word *sharik* 'balloon' was activated (Marian & Spivey, 2003a). In fact, word recognition in a non-native language is likely to be facilitated by phonological overlap with a native language (Marian et al., 2008). Moreover, studies of the effect of semantic context on word recognition report that activation of a L1 word is mitigated by semantic incongruence, where the L1 word is inconsistent with the context of the L2 sentence (C. G. Chambers & Cooke, 2009; FitzPatrick & Indefrey, 2010; Lagrou, Hartsuiker, & Duyck, 2013). For example, although the English 'pool' and the French *poule* 'chicken' are near-homophone, activation of 'pool' is weak while listening to the French sentence *Marie va nourrir la poule*, 'Marie will feed the chicken' (C. G. Chambers & Cooke, 2009). Thus, such parallel activation of phonological representations in two words might assist detection of the distributional pattern of L2 lexicons in the case of loanwords originating from English, as the borrowed words are phonologically and semantically similar to their English equivalents in general.

In addition, Hoshino and Kroll (2008) found cognate effects in a picture-naming task while L1 Spanish-L2 English and L1 Japanese-L2 English speakers were producing the name of the cognate (i.e., phonologically and semantically overlap between L1 and L2 languages) and noncognate (i.e., only semantic overlap) pictures in English. For Japanese speakers, cognate stimuli are shared phonologically and semantically, but do not orthographically overlap as Japanese and English have different scripts. For the two groups, English, Spanish, and Japanese cognates and speakers' L1-L2 cognates facilitated word production faster and more accurately than that of noncognates and non-native language-L2 cognates. Thus, irrespective of the written form of languages, cognate facilitation was observed.

Taken together, this literature suggests that L1 English-L2 Japanese learners may perform better than other language L2 speakers who learn Japanese in an experiment like the current study, which asked participants to judge the form of English loanwords; learners might have an advantage in detecting distributional patterns in English loanwords of Japanese and the patterns are primed. Therefore, an important question for future research is whether L2 learners of Japanese whose native language is other than English are able to learn Japanese loanword phonotactics/phonology. The challenge for those learners is to detect differences of sound patterns between the target language and the source language which is not their native language. Therefore, detecting distributional patterns in loanwords might be more difficult for non-English speaking L2 learners of Japanese than English-speaking L2 learners.

In addition, a complication is not knowing which listeners in the first experiment could not hear the difference between the velar singletons and geminates. Future work could assess phonological categories and phonotactics in the same learners.

Although we found that L2 acquisition of sublexicon phonology/phonotactics of a language is possible without instruction to learn in natural language settings, the performance was not like that of native Japanese speakers. The degree to which phonotactics of sublexicons was learned varied between native speakers and learners. This is a remaining question: Why are L2 learners not able to generalise the gradient geminate patterns in loanword, while native speakers of Japanese successfully learn gradient rules?

In comparison to L1 acquisition, auditory information is not the only input but a variety of knowledge sources are available for L2 acquisition (Cutler, 2015). Since a substantial number of studies have discussed the factors both constraining and promoting L2 phonological acquisition, I consider these factors at the same time as considering the ‘statistical learning’ phonotactic literature in this section to address the remaining question. First, the effect of exposure to L2 and its limitation is discussed. Second, plausible contributors on statistical L2 learning in relation to L1 acquisition in statistical language learning is discussed. Then, a plausible constraint on statistical learning is discussed by considering studies that have shown a link between statistical learning and a primacy effect.

### **6.2.2 The effect of individual degree of exposure to L2**

While this thesis focuses on the relation between L2 sublexicon phonology in Japanese and statistical language learning, it also considers the association between L2 learners’ varying degrees of exposure to the target language and their performance in the experimental tasks. In



the present study, informed quantitative measures of L2 exposure were used by applying PCA analyses to a questionnaire, in order to measure individuals' level of exposure to L2. The derived PCA factors as predictors of the well-formedness performance did not show the expected results. Specifically, we predicted L2 learners with high exposure to Japanese would be expected to be able to acquire fine-grained knowledge regarding the effects of voicing. This prediction was not borne out.

As reminder, two PCA factors were identified for L1 English-L2 Japanese speakers. The first component (PC-JJ) was related to exposure to Japanese in Japan. The second component (PC-JNZ) was related to exposure to Japanese in NZ, not derived from interpersonal communication. There are clear differences between L1 Japanese-L2 English and L1 English-L2 Japanese speakers for their L2 situations. Japanese speakers were resident in the L2 country. Therefore, a Japanese participant's degree of exposure is L2 immersion in NZ, where they need to use their L2 in their daily life. On the other hand, English-speaking learners of Japanese were studying Japanese in their native country. Thus, either exposure to Japanese in Japan or in NZ may be limited as language experience in order to detect gradient distributional patterns. Language immersion in a L2 mitigates the access to L1 (Linck et al., 2009) and increases opportunities to encounter and practice new words in L2 without extra effort in the learning (Kojic-Sabo & Lightbown, 1999).

It should be noted that individual grade levels (i.e., first-year students, second-year students) at university were not possible to investigate. In general, as learners move up through the grades, exposure to the target language increases. The sample size was small for each grade and the grade in the questionnaire was not derived by PCA. A relatively small number of participants was the most important limitation of this study. English-speaking learners of Japanese were chosen to try to control their first language background as taking care of L1 influence on the processing of L2 inputs, but this led to a small number of participants since many learners of Japanese are not native speakers of English. A future study should increase numbers and see the effects on PCA analyses, which will increase statistical power for mixed effect models.

Since the positive effect of exposure was not found, there are two possible accounts that must be considered for differences in capturing the statistics of sublexicons between native speakers and learners. As a reminder, the discussion below addresses why L2 learners are not able to generalise the gradient geminate patterns in loanwords, while native speakers of Japanese successfully learn gradient rules.

One possibility for varying the degree of performance stems from the development of L2 segmental perception. The other possibility is individuals' vocabulary size. First, the development of L2 segmental perception is discussed.

### **6.2.3 From perceptual knowledge to higher level phonological knowledge**

As discussed, a considerable amount of research indicates that language acquisition requires learners to extract regularities from inputs to which learners are exposed. The current study indicates that the phonemic level of perception of L2 speech segments is a prerequisite for detecting such regularities in natural language. As seen in the second experiment, learners are capable of perceiving singleton-geminate contrasts to some extent, even during high memory demand, but these contrasts might not yet be phonologised by some learners. That is, consonant length contrast is not phonemic and lexical (e.g., *oto* 'sound' and *otto* 'husband' in Japanese). This speculation arises from a post-hoc test that shows that L1 English-L2 Japanese listeners were significantly less accurate on geminate/singleton order than on singleton/geminate order regardless of POA or voicing types. Further research is needed, but our speculation is as follows. When listeners are responding to a geminate, they pay attention to phonetic detail and hear the geminate as 'different'. Whereas even if they hear the earlier geminate as 'different', it is phonologically stored as a singleton. By the time they come to judge the later singleton, all that listeners remember about the earlier phoneme is the category but not the detail. Even though L2 learners have meta-knowledge that obstruent consonants have a phonemic length contrast in Japanese, if learners are unable to perceive the contrasts phonemically in their daily life, they fail in mapping input to appropriate lexical distinction and in detecting statistical distributional regularities in lexicons. This might be a reason why that native speakers of Japanese successfully learn gradient rules, but L2 learners are not able to generalise the rules.

Although I acknowledge that there are similarities and differences between L1 and L2 acquisition, the findings are in line with typical phonological development in L1 acquisition that articulatory and perceptual knowledge are considered a lower level of phonological knowledge in comparison to higher level phonological knowledge, which is language-specific and gradient rather than absolute (Munson, Edwards, & Beckman, 2005). Studies analysing early speech perception have shown that 7.5-month-old infants' native contrast discrimination skills are positively associated with later language ability such as productive vocabulary size and utterance complexity (Kuhl et al., 2008; Kuhl, Conboy, Padden, Nelson, & Pruitt, 2005; Tsao, Liu, & Kuhl, 2004). Kuhl et al. (2005) propose that better speech perception facilitates

detection of phonotactic patterns in the child's native language, and such sensitivity to segmental distributional properties further assists development of native-language perception. Such linguistic development loops may apply to L2 phonological acquisition. In addition, phonologisation appears to emerge when young children become sensitive to phonetic properties of the input when they increase their vocabulary (Best & Tyler, 2007).

According to Bundgaard-Nielsen et al. (2011a), such phonological development with vocabulary growth is observed in L2 phonological development as well. For L2 learners, L2 perception is associated with their L2 vocabulary size rather than L2 exposure duration. The perception of Australian English vowels by Japanese learners after only 4–8 weeks in Australia was compared with their perception after 6–8 months of L2 exposure, in relation to their vocabulary size which was tested by a vocabulary size test (Nation & Beglar, 2007). As a result, students with larger English vocabularies (above 6,000 words) discriminated better on L2 vowel contrasts and more consistently assimilated L2 vowels to L1 categories than those with smaller vocabularies. Bundgaard-Nielsen et al. states, “L2 vocabulary expansion drives changes to L2 segmental perception, which reflects better reattunement and rephonologization” (p. 447). Importantly, this study showed that the length of exposure (i.e., increased language experience of L2) is not an influential factor on improvement of L2 vowel perception. This leads to another potential possibility to account for differences in statistical learning between Japanese speakers and L2 learners.

In sum, the differences in gradient rule learning between native Japanese speakers and learners might be due to that L2 learners have not achieved higher levels of phonological knowledge in L2 in detecting gradient distributional patterns. Although they are able to discriminate the L2 consonantal contrasts phonetically in the perception task to some degree, at least some participants are unable to phonologise them. Therefore, L2 learners do not acquire phonotactic distributional patterns of consonant gemination in loanwords. This speculation is an issue for a future study.

#### **6.2.4 Influence of vocabulary size on statistical learning**

Another possibility is that varying degrees of performance in the well-formedness task are connected to quantitative aspects of lexical knowledge, as discussed in §1.2.2 and §2.3.3.1. That is, vocabulary size might play a significant role in facilitating detection of distribution information in input. In such an account, gradient statistical knowledge increases due to the increased size of vocabulary, as it contains various exemplars that assist in detecting statistical

distributional patterns in sublexicons. The average vocabulary of university students in Japan generally consists of around 39,000-40,000 words (Sato et al., 2017). Based on the analyses of the BCCWJ in Chapter 2, if they have such vocabulary size, the estimated loanword vocabulary size of native Japanese speakers is approximately 4,300 words. As for learners, they tend to have only about 300 loanwords, if their vocabulary consists of around 3,000 to 4,000 Japanese words. However, this estimation does not apply to some learners in the current study, who had just completed the first level of their Japanese course. In order to detect gradient distributional patterns, L2 learners might need to acquire a certain number of word types in their vocabulary (or reach above a threshold of vocabulary size), providing learners with sufficient data to detect statistical distributional patterns across sublexicons.

Studies about acquiring new languages after short exposure indicate that having a rich lexicon is not necessary to detect statistical distributional patterns at the outset of learning. For example, in an experimental task with adult Dutch speakers, Gullberg, Roberts, Dimroth, Veroude, and Indefrey (2010) found that even after short exposure to a new language (Mandarin Chinese), naïve learners were able to detect syllable structure violations in the language. On the other hand, the current study suggests that vocabulary size appears to be potentially important for acquiring gradient phonotactic knowledge. This perspective does not create a contradiction when considering statistical learning on L1 acquisition. Many studies show that infants exhibit early sensitivity to possible sound patterns in their native languages before they begin producing words (e.g., Jusczyk et al., 1993; Jusczyk et al., 1994; Kuhl et al., 2005; Mattys & Jusczyk, 2001; Mattys et al., 1999), which may not require ample vocabulary. Various studies have examined the relation between young children's phonological awareness and vocabulary growth, as they develop linguistic experience (e.g., Edwards et al., 2004; Graf Estes et al., 2016; Metsala, 1999; Stokes, Moran, & George, 2013; Storkel, 2001). The 18-month-infants who possessed small vocabularies showed greater flexibility in learning novel phonotactic patterns that were illicit in their native language sound patterns, but similarly aged infants with medium- and large-sized vocabularies did not (Graf Estes et al., 2016). Edwards et al. (2004) showed that during a nonword repetition task attempted by 3-8 year-old children, raw measures of vocabulary size was the best predictor of overall accuracy and the effect of sequence frequency on accuracy than their age. Children with larger vocabularies repeated low-frequency and zero-frequency clusters more accurately than young children with smaller vocabularies. Thus, children with larger vocabularies tend to have a more mature higher-level phonological knowledge. Graf Estes et al. (2016, p. 13) state that, “[a]s vocabulary knowledge is stored, learners gather rich information supporting generalizations about how frequently

sounds occur in the input, how frequently sounds occur together, and where those sound and sound combinations occur within words”. Such trajectory of phonological development seems to be required for L2 learners.

With respect to adults, Frisch et al. (2001) found that individuals with a larger vocabulary are more likely to accept low probability English nonwords than individuals with a smaller vocabulary. This indicates that individuals with greater lexical knowledge are less likely to treat low probability nonwords uniformly, than that of individuals with less lexical knowledge. Frisch et al. proposed that the threshold of acceptability for nonwords is inversely related to vocabulary size. This is based on the assumption that a larger vocabulary would provide individuals with more exposure to less frequent phonotactic patterns in a language. Frisch and Brea-Spahn (2010) confirmed the associative relation between well-formedness judgements and lexical knowledge in adult English monolinguals and Spanish-English bilinguals, when they judged well-formedness for English nonwords with onset-rime phonotactic probabilities. This suggests that the judgments of bilingual speakers are similar to that of monolingual speakers. Regardless of their first language, participants with larger vocabularies in English were more likely to accept low probability nonwords in English than participants with smaller vocabularies. It should be noted that the vocabulary effects were not found in a Spanish nonword task in which only bilinguals judged Spanish nonwords. However, importantly vocabulary effects on well-formedness judgments within English suggest that lexical judgement would be gradient with the increase of lexical knowledge. In the current study, although voiced geminates were treated by L2 learners as if these geminates occur equally frequently, sensitivity to high/low-probability phonotactic patterns might be increased by learners increasing their vocabulary size. In fact, few examples with voiced labial geminates in loanwords were found in the BCCWJ corpus. Since L2 learners know fewer words than native speakers, it is less likely for L2 learners to encounter such words.

In sum, these studies indicate the importance of vocabulary size to detect gradient distributional patterns in lexicons, as gradient knowledge reflects type and token frequency in the input. Sufficient quantity and quality of input are needed for statistical learning on gradient phonotactics to happen during acquisition of sublexicon phonology. However, recall that the Japanese vocabulary size for participants was not directly measured in the present study. It was estimated based on their expected overall vocabulary size by using frequency information in the BCCWJ corpus, which was not sensitive in measuring lexical knowledge in comparison to assessing individual receptive vocabulary size. There was no direct investigation of whether statistical learning of sublexicon phonology is associated with individuals’ L2 vocabulary size.

This is another limitation of the present study. Future studies should examine L2 lexical knowledge using a vocabulary size test as an indicative measure. Such measurements would enhance a more precise understanding of the development of learning gradient phonological knowledge and might provide evidence of the importance of the vocabulary size on L2 statistical language learning. In addition, such measurements may enable us to better understand the great variability in individual perceptual discrimination on non-native sound contrasts.

### **6.2.5 Possible constraints for statistical learning**

Lastly, this section discusses a possible constraint on statistical learning. In a review of statistical learning by Aslin (2017), he describes four types of constraints on statistical learning: (1) attention, (2) perceptual biases, (3) prosody and (4) primacy and familiarity. Although all of them are possible factors to constrain the statistical learning of sublexicon phonology for language users, I discuss ‘primacy effect’ in relation to ‘overlearning’ briefly. Several studies report that the primacy effect, by which adults learn initial statistical structure but not second statistical structure without contextual cues such as a speaker’s voice, when two successive conflicting artificial structures were presented (Gebhart, Aslin, & Newport, 2009; Mitchel & Weiss, 2010; Weiss, Gerfen, & Mitchel, 2009). On the other hand, Bulgarelli and Weiss (2016) suggest that only learning the first structure of two inputs is presumably due to overlearning.

The results of the well-formedness judgment task indicate overgeneralising the more frequent patterns in loanwords, as discussed shortly in §4.5.2. Although wrong epenthesis in alveolars (e.g., /bettu/) are rated significantly lower than singletons (e.g., /beto/) and geminates (e.g., /betto/) in the same contexts, they were rated significantly higher than wrong epenthesis in labial (e.g., /pippo/) and velar (e.g., /pakko/) contexts. For the quality of epenthetic vowel, the epenthetic vowel [u] has a higher frequency as it is used after 10 coda consonants /p, b, k, g, ɸ, s, ʃ, z, m, r/ in Japanese when foreign words are borrowed in Japanese. On the other hand, the vowel [o] occurs only after alveolar stops /t, d/. Thus, it is possible that language users overgeneralise an epenthetic rule in which [u] can be used in any context. For geminates, voiced geminates are less likely than voiceless geminates, even in loanwords. Therefore, learners are more likely to encounter voiceless geminates in our target instances (i.e., adaptation of English final consonants). As discussed in connection to rule learning, in §1.3, the likelihood of structural patterns in the input enables learners make a broader generalisation (i.e., AAB/ABA), or a narrow generalisation (i.e., AAdi/AdiA) (Gerken, 2006). In addition, when adult learners

were exposed to the stimulus sets containing incomplete overlap used in the previous experiment thrice rather than once, they are less likely to generalise to new sentences (Reeder et al., 2013). That is, a distributional pattern detected by learners facilitates detection of similar instances, making phonological generalisation to novel exemplars, but hindering the detection of a different pattern. Bulgarelli and Weiss (2016) argue that after learners have achieved robust learning of their first structure, they are less attentive to a second structure. In such an account, for both epenthetic vowel and geminates, the frequent observed regularities may block learners from acquiring less frequent distributional patterns. Once learners acquire a first structure (e.g., epenthetic [u]) that works after most consonants, they then apply this generalisation to all instances, which is difficult to unlearn.

Last of all, there is an increasing amount of literature that discusses individual differences in statistical learning (Siegelman, Bogaerts, Christiansen, & Frost, 2017; Siegelman & Frost, 2015), some studies finding that adult individual differences in statistical learning are related to language outcome (e.g., Frost, Siegelman, Narkiss, & Afek, 2013; Kaufman et al., 2010; Misyak & Christiansen, 2007; Potter et al., 2017). For example, Kaufman et al. (2010) found correlation between adult implicit learning ability measured by serial reaction time tasks and performance in two foreign language exams. Similarly, English speaking learners of Hebrew who detected embedded statistical structure better in a visual statistical learning task were more likely to succeed in assimilating Hebrew word morphology (Frost et al., 2013). Misyak and Christiansen (2012) also found individual's statistical learning performance on learning of artificial syntactic grammar tasks (i.e., adjacent and nonadjacent regularities) was strongly interrelated with verbal working memory and language comprehension, among other language relevant factors. Furthermore, statistical learning performance on the two regularities predicted processing ability for two types of sentences involving local or long-distance dependencies, rather than verbal working memory. In addition, Potter et al. (2017) found that L2 experience in introductory Mandarin classes facilitated learners' performance at artificial tonal statistical learning, but not in relation to visual statistical learning.

Thus, in general, an increased ability to detect the distributional patterns of non-linguistic inputs in the environment predicts that learners are more effective in detecting a new set of statistical regularities in languages, but linguistically-relevant experience influences only domain-relevant statistical learning. Such an approach might help to account for differences in performance between learners, but not differences between native Japanese speakers and L2 learners in their performance in the current study. Additionally, Kaufman et al. (2010) found

that implicit learning was strongly related to self-reported personality, including intuition. Such cognitive ability could be related to statistical language learning of a sublexicon phonology but it is beyond the scope of this thesis.

### **6.3 Conclusion**

This thesis has explored whether L2 learners are able to intuit and detect statistical properties of sound patterns in sublexicons as generated from their entire L2 lexicons. Using loanword phonology and phonotactics in Japanese as a case study, an influence of sublexicons in Japanese on well-formedness judgments was demonstrated. L2 learners exhibit greater sensitivity to the sublexicon-specific phonotactics (i.e., Japanese loanword phonotactics) rather than overall Japanese phonotactics on perceived lexical well-formedness judgments. This is attributed to the underlying statistics based on accumulated knowledge which learners acquire from instances of natural language they are exposed to, without being taught. Thus, the results show that a powerful statistical learning mechanism is used in L2 language acquisition of the sublexicon phonology. The findings in this thesis help clarify that the learnability of sublexicon properties in L2 language acquisition is underpinned by statistical knowledge of loanwords to which learners are exposed. Importantly, the capacity for statistical learning is not restricted to overall phonotactics in a language. The findings extend the statistical learning literature, which has typically concentrated on the learning of overall phonotactics in a language.

However, the present results suggest that categorical learning, such as the quality of epenthetic vowels, was possible even though language users have small vocabulary size, but not in regard to detecting the gradient stochastic distributional patterns of voiced geminates (i.e., gradient rules). These findings suggest that categorical rules have been more easily learned than gradient ones, when learners possess a small vocabulary. Since native speakers of Japanese show greater sensitivity to gradient distributional patterns, as vocabulary increases, so does their sensitivity to gradient stochastic patterns.

The memory decision task showed that native speakers demonstrated a phonemic level of processing on native contrastive sounds, whereas L2 learners may perform a phonetic level of processing on non-native contrastive sounds. That is, learners are able to discriminate non-native contrastive sounds (i.e., singleton/geminate) phonetically to some extent, but they may be unable to phonologise these sounds as yet. This is speculated from a post-hoc test showing that L1 English-L2 Japanese listeners were significantly less accurate on geminate/singleton order than on singleton/geminate order regardless of POA or voicing types. This suggests that they



can 'hear' the long consonant which affects their responses, if responding immediately. But they do not store it as separately from the singleton. This might be a reason why that L2 learners are not able to generalise gradient rules.

Together, with the results of the two experiments, these findings suggest that gradient phonological knowledge increases along with vocabulary growth, because different types of phoneme sequences language users are exposed to also increase. In addition, L2 segmental perception would be also improved by L2 vocabulary development.

Natural languages are complex, especially as different phonological systems coexist within a single language. The study of a sublexicon phonology in Japanese provide insights into both the general understanding of phonological aspects of language acquisition and L2 phonological awareness. This thesis demonstrated that not only native speakers of a language but also L2 learners are sensitive to the phonologically multidimensional structure of a natural language. Language learners can implicitly detect the statistical structure of a sublexicon phonology in a language over the course of acquiring a natural language.

# References

- Akamatsu, T. (2000). *Japanese phonology: A functional approach* (Vol. 38). Munich, Germany: Lincom Europa.
- Albright, A. (2009). Feature-based generalisation as a source of gradient acceptability. *Phonology*, 26(1), 9-41.
- Altenberg, E. P. (2005). The judgment, perception, and production of consonant clusters in a second language. *IRAL-International Review of Applied Linguistics in Language Teaching*, 43(1), 53-80.
- Altmann, H., Berger, I., & Braun, B. (2012). Asymmetries in the perception of non-native consonantal and vocalic length contrasts. *Second Language Research*, 28(4), 387-413. doi:0.1177/0267658312456544
- Amano, N., & Kondo, K. (2000). *Nihongo-no goi tokusei* [Lexical properties of Japanese]. Tokyo: Sanseido.
- Aoyama, K. (2003). Perception of syllable-initial and syllable-final nasals in English by Korean and Japanese speakers. *Second Language Research*, 19(3), 251-265. doi:10.1191/0267658303sr222oa
- Asano, Y. (2018). Discriminating non-native segmental length contrasts under increased task demands. *Language and Speech*, 61(3), 409-429. doi:10.1177/0023830917731907
- Aslin, R. N. (2017). Statistical learning: a powerful mechanism that operates by mere exposure. *Wiley Interdisciplinary Reviews: Cognitive Science*, 8(1-2), 1-7. doi:10.1002/wcs.1373
- Aslin, R. N., & Newport, E. L. (2012). Statistical learning: From acquiring specific items to forming general rules. *Current Directions in Psychological Science*, 21(3), 170-176.
- Aslin, R. N., Saffran, J. R., & Newport, E. L. (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychological Science*, 9(4), 321-324.
- 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. doi:10.1016/j.jml.2007.12.005
- Baayen, R. H., Piepenbrock, R., & Van Rijn, H. (1993). The CELEX lexical database (CD-ROM). Linguistic data consortium. Philadelphia, PA: University of Pennsylvania.
- Bailey, T. M., & Hahn, U. (2001). Determinants of wordlikeness: Phonotactics or lexical neighborhoods? *Journal of Memory and Language*, 44(4), 568-591.

- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3), 255-278. doi:10.1016/j.jml.2012.11.001
- Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., . . . Bolker, M. B. (2015). Package ‘lme4’. *Convergence*, 12(1), 2.
- Beavers, A. S., Lounsbury, J. W., Richards, J. K., & Huck, S. W. (2013). Practical considerations for using exploratory factor analysis in educational research. *Practical Assessment, Research, and Evaluation*, 18(1), 6.
- Beckman, M. E. (1982). Segment duration and the ‘mora’ in Japanese. *Phonetica*, 39(2-3), 113-135.
- Berent, I., Steriade, D., Lennertz, T., & Vaknin, V. (2006). What we know about what we have never heard: evidence from perceptual illusions. *Cognition*, 104(3), 591-630. doi:10.1016/j.cognition.2006.05.015
- Best, C. T. (1994). The emergence of native-language phonological influences in infants: A perceptual assimilation model. In J. Goodman & H. Nusbaum (Eds.), *The development of speech perception: The transition from speech sounds to spoken words* (pp. 167-224). Cambridge: MIT Press.
- Best, C. T. (1995). A direct realist view of cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language speech research* (Vol. 92, pp. 171-204). Timonium, MD: York Press.
- Best, C. T., McRoberts, G. W., & Goodell, E. (2001). Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener’s native phonological system. *Journal of the Acoustical Society of America*, 109(2), 775-794. doi:10.1121/1.1332378
- Best, C. T., & Tyler, M. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In O.-S. Bohn & M. Munro (Eds.), *Language experience in second language speech learning: In honour of James Emil Flege* (pp. 13-34). Amsterdam, Philadelphia: John Benjamins.
- Bice, K., & Kroll, J. F. (2015). Native language change during early stages of second language learning. *NeuroReport*, 26(16), 966. doi:10.1097/WNR.0000000000000453
- Boersma, P., & Weenink, D. (2018). Praat: Doing phonetics by computer [Computer program]. (Version 6.0. 37).

- Bradlow, A. R., Nygaard, L. C., & Pisoni, D. B. (1999). Effects of talker, rate, and amplitude variation on recognition memory for spoken words. *Perception & Psychophysics* 61(2), 206-219.
- Breiman, L. (2001). Random forests. *Machine learning*, 45(1), 5-32.
- Bulgarelli, F., & Weiss, D. J. (2016). Anchors aweigh: The impact of overlearning on entrenchment effects in statistical learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 42(10), 1621-1631. doi:10.1037/xlm0000263
- Bundgaard-Nielsen, R. L., Best, C. T., & Tyler, M. D. (2011a). Vocabulary size is associated with second-language vowel perception performance in adult learners. *Studies in Second Language Acquisition*, 433-461. doi:10.1017/S0272263111000040
- Bundgaard-Nielsen, R. L., Best, C. T., & Tyler, M. D. (2011b). Vocabulary size matters: The assimilation of second-language Australian English vowels to first-language Japanese vowel categories. *Applied Psycholinguistics*, 32(1), 51.
- Byarushengo, E. R. (1976). *Strategies in loan phonology*. Paper presented at the Annual Meeting of the Berkeley Linguistics Society.
- Campbell, N. (1992). Segmental elasticity and timing in Japanese speech. In Y. Tokuhara, E. V. Vatikiotis-Bateson, & Y. Sagisaka (Eds.), *Speech perception, production and linguistic structure* (pp. 403-418). Tokyo: Ohmsha.
- Chambers, C. G., & Cooke, H. (2009). Lexical competition during second-language listening: Sentence context, but not proficiency, constrains interference from the native lexicon. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35(4), 1029.
- Chambers, K. E., Onishi, K. H., & Fisher, C. (2003). Infants learn phonotactic regularities from brief auditory experience. *Cognition*, 87(2), B69-B77.
- Chambers, K. E., Onishi, K. H., & Fisher, C. (2011). Representations for phonotactic learning in infancy. *Language Learning and Development*, 7(4), 287-308.
- Chan, C. L. (2018). Speech In Noise 2 [Computer program]: Northwestern University. Retrieved from <https://northwestern.app.box.com/s/9g2rigz1iqh4ymfkgunq6t31u3iycbpr>
- Chang, C. B. (2012). Rapid and multifaceted effects of second-language learning on first-language speech production. *Journal of Phonetics*, 40(2), 249-268. doi:10.1016/j.wocn.2011.10.007
- Chang, C. B. (2013). A novelty effect in phonetic drift of the native language. *Journal of Phonetics*, 41(6), 520-533. doi:10.1016/j.wocn.2013.09.006

- Chen, J., Best, C. T., & Antoniou, M. (2019). Cognitive factors in Thai-naïve Mandarin speakers' imitation of Thai lexical tones. *Crossroads of Speech and Language: Proceedings of INTERSPEECH 2019, 15-19 September 2019, Graz, Austria*, 2653-2657.
- Coleman, J., & Pierrehumbert, J. B. (1997). *Stochastic phonological grammars and acceptability*. Paper presented at the Third Meeting of the ACL Special Interest Group in Computational Phonology: Proceedings of the Workshop.
- Cowan, N., & Morse, P. A. (1986). The use of auditory and phonetic memory in vowel discrimination. *Journal of the Acoustical Society of America*, 79(2), 500-507. doi:10.1121/1.393537
- Crawford, C. (2009). *Adaptation and transmission in Japanese loanword phonology*. Doctoral dissertation, Cornell University.
- Crowder, R. G. (1982). A common basis for auditory sensory storage in perception and immediate memory. *Perception & Psychophysics*, 31(5), 447-483.
- Cutler, A. (2015). Representation of second language phonology. *Applied Psycholinguistics*, 36(1), 115-128.
- Darcy, I., Mora, J. C., & Daidone, D. (2014). Attention control and inhibition influence phonological development in a second language. *Concordia Working Papers in Applied Linguistics*, 5, 115-129.
- Davidson, L. (2006). Phonology, phonetics, or frequency: Influences on the production of non-native sequences. *Journal of Phonetics*, 34(1), 104-137.
- Davidson, L. (2010). Phonetic bases of similarities in cross-language production: Evidence from English and Catalan. *Journal of Phonetics*, 38(2), 272-288
- Dehaene-Lambertz, G., Dupoux, E., & Gut, A. (2000). Electrophysiological Correlates of Phonological Processing: A Cross-linguistic Study. *Journal of Cognitive Neuroscience*, 12(4), 635-647. doi:10.1162/089892900562390
- Dunteman, G. H. (1989). *Principal components analysis*: Sage.
- Dupoux, E., Kakehi, K., Hirose, Y., Pallier, C., & Mehler, J. (1999). Epenthetic vowels in Japanese: A perceptual illusion? *Journal of Experimental Psychology: Human Perception and Performance*, 25(6), 1568-1578. doi:10.1037/0096-1523.25.6.1568
- Dupoux, E., Parlato, E., Frota, S., Hirose, Y., & Peperkamp, S. (2011). Where do illusory vowels come from? *Journal of Memory and Language*, 64(3), 199-210. doi:10.1016/j.jml.2010.12.004

- Edwards, J., Beckman, M. E., & Munson, B. (2004). The interaction between vocabulary size and phonotactic probability effects on children's production accuracy and fluency in nonword repetition. *Journal of Speech, Language, and Hearing Research*, 47, 421–436.
- Ernestus, M. (2011). Gradience and categoricity in phonological theory. In M. van Oostendorp, C. Ewen, E. Hume, & K. Rice (Eds.), *The Blackwell companion to phonology* (pp. 2115–2136). Malden, MA: Wiley Blackwell.
- Field, A., Miles, J., & Field, Z. (2012). *Discovering statistics using R*. London: Sage.
- First People's Cultural Council. (2013). *Template 1: Language Assessment Survey - First Peoples*. Retrieved from <https://www.yumpu.com/en/document/read/52040422/template-1-language-assessment-survey-first-peoples>
- FitzPatrick, I., & Indefrey, P. (2010). Lexical competition in nonnative speech comprehension. *Journal of Cognitive Neuroscience*, 22(6), 1165-1178.
- Flege, J. E. (1991). Perception and production: The relevance of phonetic input to L2 phonological learning. In C. Ferguson & T. Heubner (Eds.), *Crosscurrents in second language acquisition and linguistic theories* (Vol. 2, pp. 249-289). Philadelphia: J Benjamin.
- Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 299-273). Timonium, MD: York Press.
- Flege, J. E. (1999). Age of learning and second language speech. In D. Birdsong (Ed.), *Second language acquisition and the critical period hypothesis* (pp. 101-131). Mahwah, NJ: Erlbaum.
- Flege, J. E. (2009). Give input a chance. In M. Young-Scholten & T. Piske (Eds.), *Input matters in SLA* (pp. 175-190). Clevedon: Multilingual Matters.
- Flege, J. E., Schirru, C., & MacKay, I. R. (2003). Interaction between the native and second language phonetic subsystems. *Speech Communication* 40(4), 467-491.  
doi:10.1016/S0167-6393(02)00128-0
- Flege, J. E., Yeni-Komshian, G. H., & Liu, S. (1999). Age constraints on second-language acquisition. *Journal of Memory and Language*, 41(1), 78-104.
- Fleischhacker, H. (2001). Cluster-dependent epenthesis asymmetries. *UCLA working papers in linguistics*, 7, 71-116.

- Frauenfelder, U., Segui, J., & Dijkstra, T. (1990). Lexical effects in phonemic processing: Facilitating or inhibitory? *Journal of Experimental Psychology: Human Perception and Performance*, *16*(1), 77-91.
- Friederici, A. D., & Wessels, J. M. (1993). Phonotactic knowledge of word boundaries and its use in infant speech perception. *Perception & Psychophysics*, *54*(3), 287-295.
- Frisch, S. A., & Brea-Spahn, M. R. (2010). Metalinguistic judgments of phonotactics by monolinguals and bilinguals. *Laboratory Phonology*, *1*(2), 345-360.
- Frisch, S. A., Large, N. R., & Pisoni, D. B. (2000). Perception of wordlikeness: Effects of segment probability and length on the processing of nonwords. *Journal of Memory and Language*, *42*(4), 481-496.
- Frisch, S. A., Large, N. R., Zawaydeh, B., & Pisoni, D. B. (2001). Emergent phonotactic generalizations in English and Arabic. In J. Bybee & P. J. Hopper (Eds.), *Frequency and the emergence of linguistic structure* (Vol. 45, pp. 159-180). Amsterdam: Benjamins.
- Frost, R., Siegelman, N., Narkiss, A., & Afek, L. (2013). What predicts successful literacy acquisition in a second language? *Psychological Science*, *24*(7), 1243-1252.
- Ganong, W. F. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance*, *6*(1), 110-125.  
doi:10.1037/0096-1523.6.1.110
- Gebhart, A. L., Aslin, R. N., & Newport, E. L. (2009). Changing Structures in Midstream: Learning Along the Statistical Garden Path. *Cognitive Science*, *33*(6), 1087-1116.  
doi:10.1111/j.1551-6709.2009.01041.x
- Gelbart, B. (2005). *Perception of foreignness*. Doctoral dissertation, University of Massachusetts Amherst.
- Gelbart, B., & Kawahara, S. (2007). *Lexical cues to foreignness in Japanese*. Paper presented at the Proceedings of Formal Approaches to Japanese Linguistics.
- Gerken, L. (2006). Decisions, decisions: Infant language learning when multiple generalizations are possible. *Cognition*, *98*(3), B67-B74.
- Gerken, L. (2010). Infants use rational decision criteria for choosing among models of their input. *Cognition*, *115*(2), 362-366.
- Gerrits, E., & Schouten, M. E. H. (2004). Categorical perception depends on the discrimination task. *Perception & Psychophysics*, *66*(3), 363-376.

- Gibson, E., Piantadosi, S., & Fedorenko, K. (2011). Using Mechanical Turk to Obtain and Analyze English Acceptability Judgments. *Language and Linguistics Compass*, 5(8), 509-524. doi:10.1111/j.1749-818X.2011.00295.x
- Gómez, R. L., & Gerken, L. (1999). Artificial grammar learning by 1-year-olds leads to specific and abstract knowledge. *Cognition*, 70(2), 109-135.
- Gómez, R. L., & Lakusta, L. (2004). A first step in form-based category abstraction by 12-month-old infants. *Developmental Science*, 7(5), 567-580.
- Graf Estes, K., Edwards, J., & Saffran, J. R. (2011). Phonotactic constraints on infant word learning. *Infancy*, 16(2), 180-197.
- Graf Estes, K., Gluck, S. C.-W., & Grimm, K. J. (2016). Finding patterns and learning words: Infant phonotactic knowledge is associated with vocabulary size. *Journal of Experimental Child Psychology*, 146, 34-49.
- Green, D. M., & Swets, J. A. (1966). Signal detection theory and psychophysics. New York: Wiley
- Guion, S. G., & Pederson, E. (2007). Investigating the Role of Attention in Phonetic Learning. In O.-S. Bohn & M. Munro (Eds.), *Language experience in second language speech learning: in honor of James Emil Flege* (pp. 57-77). Amsterdam, Philadelphia: John Benjamins.
- Gullberg, M., Roberts, L., Dimroth, C., Veroude, K., & Indefrey, P. (2010). Adult language learning after minimal exposure to an unknown natural language. *Language Learning*, 60, 5-24.
- Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (1998). *Multivariate data analysis with Readings, 5th Edition* (Vol. 5). Englewood Cliffs, NJ: Prentice Hall.
- Hall, K. C. (2009). *A probabilistic model of phonological relationships from contrast to allophony*. Doctoral dissertation, The Ohio State University.
- Hall, N. (2011). Vowel epenthesis. In M. van Oostendorp, C. Ewen, E. Hume, & K. Rice (Eds.), *The Blackwell companion to phonology* (Vol. 3, pp. 1576-1596). Malden, MA: Wiley-Blackwell.
- Hallé, P. A., Segui, J., Frauenfelder, U., & Meunier, C. (1998). Processing of illegal consonant clusters: A case of perceptual assimilation? *Journal of Experimental Psychology: Human Perception and Performance*, 24(2), 592.
- Han, M. S. (1962). The feature of duration in Japanese. *Study Sounds*, 10, 65-80.



- Han, M. S. (1992). The timing control of geminate and single stop consonants in Japanese: A challenge for nonnative speakers. *Phonetica*, 49(2), 102-127. doi:10.1159/000261906
- Han, M. S. (1994). Acoustic manifestations of mora timing in Japanese. *Journal of the Acoustical Society of America*, 96(1), 73-82. doi:10.1121/1.410376
- Hardison, D. M., & Saigo, M. M. (2010). Development of perception of second language Japanese geminates: Role of duration, sonority, and segmentation strategy. *Applied Psycholinguistics*, 31(1), 81-99. doi:10.1017/s0142716409990178
- Hashimoto, D. (2019). *Loanword phonology in New Zealand English: Exemplar activation and message predictability*. Doctoral dissertation, University of Canterbury.
- Hay, J., Pierrehumbert, J. B., & Beckman, M. E. (2004). Speech perception, well-formedness, and the statistics of the lexicon. In J. Local, R. Ogden, R. Temple, M. E. Beckman, & J. Kingston (Eds.), *Phonetic Interpretation: Papers in Laboratory Phonology VI* (pp. 58-74). Cambridge, UK: Cambridge University Press.
- Hayes, R. L. (2001). *The perception of novel phoneme contrasts in a second language: A developmental study of native speakers of English learning Japanese singleton and geminate consonant contrasts*. University of Arizona Coyote Papers: Working Papers in Linguistics, Language in Cognitive Science. University of Arizona. Tucson.
- Hendrix, P., Bolger, P., & Baayen, R. H. (2017). Distinct ERP signatures of word frequency, phrase frequency, and prototypicality in speech production. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 43(1), 128-149. doi:10.1037/a0040332
- Henry, L. (2011). *The development of working memory in children*. London: Sage.
- Hirata, Y. (2004). Training native English speakers to perceive Japanese length contrasts in word versus sentence contexts. *Journal of the Acoustical Society of America*, 116(4 Pt 1), 2384-2394. doi:10.1121/1.1783351
- Hirata, Y. (2007). Durational Variability and Invariance in Japanese Stop Quantity Distinction: Roles of Adjacent Vowels. *Journal of the Phonetic Society of Japan*, 11(1), 9-22. doi:10.24467/onseikenkyu.11.1\_9
- Hirata, Y., Whitehurst, E., & Cullings, E. (2007). Training native English speakers to identify Japanese vowel length contrast with sentences at varied speaking rates. *Journal of the Acoustical Society of America*, 121(6), 3837-3845. doi:10.1121/1.2734401
- Hirata, Y., & Whiton, J. (2005). Effects of speaking rate on the single/geminate stop distinction in Japanese. *Journal of the Acoustical Society of America*, 118(3), 1647-1660. doi:10.1121/1.2000807

- Hirayama, M. (2003). Contrast in Japanese vowels. *Toronto Working Papers in Linguistics*, 20, 115-132.
- Hirayama, M. (2005). Place asymmetry and markedness of labials in Japanese. *Toronto Working Papers in Linguistics*, 24, 125-168.
- Hirose, A., & Ashby, M. (2007). *An acoustic study of devoicing of the geminate obstruents in Japanese*. Paper presented at the Proceedings of ICPhS.
- Hisagi, M., & Strange, W. (2011). Perception of Japanese temporally-cued contrasts by American English listeners. *Language and Speech*, 54(2), 241-264.  
doi:10.1177/0023830910397499
- Hock, H. H., & Joseph, B. D. (1996). *Language history, language change, and language relationship: An Introduction to historical and comparative linguistics* (Vol. 93). Berlin: Mouton de Gruyter.
- Homma, Y. (1981). Durational relationship between Japanese stops and vowels. *Journal of Phonetics*, 9(3), 273-281. doi:10.1016/S0095-4470(19)30971-4
- Hoshino, N., & Kroll, J. F. (2008). Cognate effects in picture naming: Does cross-language activation survive a change of script? *Cognition*, 106(1), 501-511.  
doi:doi.org/10.1016/j.cognition.2007.02.001
- Idemaru, K., & Guion, S. G. (2008). Acoustic covariants of length contrast in Japanese stops. *Journal of the International Phonetic Association*, 38(2), 167-186.
- Idemaru, K., & Guion-Anderson, S. (2010). Relational timing in the production and perception of Japanese singleton and geminate stops. *Phonetica*, 67(1-2), 25-46.  
doi:10.1159/000319377
- Ingram, J. C., & Park, S. G. (1997). Cross-language vowel perception and production by Japanese and Korean learners of English. *Journal of Phonetics*, 25(3), 343-370.
- Irwin, M. (2011). *Loanwords in Japanese*. Philadelphia, Netherlands: Benjamins.
- Isaacs, T., & Trofimovich, P. (2011). Phonological memory, attention control, and musical ability: Effects of individual differences on rater judgments of second language speech. *Applied Psycholinguistics*, 32(1), 113-140. doi:10.1017/S0142716410000317
- Ishida, M., & Arai, T. (2015). *Perception of an existing and non-existing L2 English phoneme behind noise by Japanese native speakers*. Paper presented at the Sixteenth Annual Conference of the International Speech Communication Association.
- Itô, J. (1989). A prosodic theory of epenthesis. *Natural Language & Linguistic Theory*, 7(2), 217-259.

- Itô, J., Kubozono, H., & Mester, A. (2017). A prosodic account of consonant gemination in Japanese loanwords. In H. Kubozono (Ed.), *The phonetics and phonology of geminate consonants* (pp. 283-320). Oxford: Oxford University Press.
- Itô, J., & Mester, A. (1995). Japanese phonology: constraint domains and structure preservation. In J. Goldsmith (Ed.), *The handbook of phonological theory* (pp. 816-838): Oxford: Blackwell.
- Itô, J. & A. Mester. (1996). Stem and Word in Sino-Japanese. In T. Otake & A. Cutler (Eds.), *Phonological Structure and Language Processing: Cross-Linguistic Studies* (pp. 13–44):New York: Mouton de Gruyter.
- Itô, J., & Mester, A. (1999). The phonological lexicon. In N. Tsujimura (Ed.), *The hand book of Japanese linguistics* (pp. 62-100). Malden, Mass, Oxford: Blackwell.
- Itô, J., & Mester, A. (2001). Covert generalizations in Optimality Theory: the role of stratal faithfulness constraints. *Studies in Phonetics, Phonology, and Morphology*, 7(2), 273-299.
- Itô, J., & Mester, A. (2015). Sino-Japanese phonology. In H. Kubozono (Ed.), *Handbook of Japanese phonetics and phonology* (pp. 289–312). Berlin: De Gruyter Mouton.
- Itô, J., Mester, A., & Padgett, J. (1995). Licensing and underspecification in Optimality Theory. *Linguistic Inquiry*, 26, 571-613.
- Ivanova, I., & Costa, A. (2008). Does bilingualism hamper lexical access in speech production? *Acta Psychologica*, 127(2), 277-288. doi:10.1016/j.actpsy.2007.06.003
- Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, 59(4), 434-446.
- Jaeger, T. F. (2011). Post to the R-LANG mailing list, 20 February 2011. Retrieved from <http://pidgin.ucsd.edu/pipermail/r-lang/2011-February/000225.html>
- Johnson, K. (2004). *Cross-linguistic perceptual differences emerge from the lexicon*. Paper presented at the the 2003 Texas Linguistics Society Conference: Coarticulation in speech production and perception.
- Johnson, K., & Martin, J. (2001). Acoustic vowel reduction in Creek: Effects of distinctive length and position in the word. *Phonetica*, 58(1-2), 81-102.
- Jusczyk, P. W., Friederici, A. D., Wessels, J. M., Svenkerud, V. Y., & Jusczyk, A. M. (1993). Infants' sensitivity to the sound patterns of native language words. *Journal of Memory and Language*, 32(3), 402-420.
- Jusczyk, P. W., Luce, P. A., & Jan, C.-L. (1994). Infants' sensitivity to phonotactic patterns in the native language. *Journal of Memory and Language*, 33(5), 630-645.

- Kabak, B., & Idsardi, W. J. (2007). Perceptual distortions in the adaptation of English consonant clusters: Syllable structure or consonantal contact constraints? *Language and Speech*, 50, 23-52. doi:10.1177/00238309070500010201
- Kaiki, N., Takeda, K., & Sagisaka, Y. (1990). *Statistical analysis for segmental duration rules in Japanese speech synthesis*. Paper presented at the First International Conference on Spoken Language Processing.
- Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika*, 39, 31-36.
- Kaneko, E., & Iverson, G. K. (2009). *Phonetic and other factors in Japanese on-line adaptation of English final consonants*. Paper presented at the eighth annual conference of the Japanese Society for Language Science Studies.
- Kang, Y. (2003). Perceptual similarity in loanword adaptation: English postvocalic word-final stops in Korean. *Phonology*, 20(02), 219-273.
- Kang, Y. (2011). Loanword phonology. In M. van Oostendorp, C. Ewen, E. Hume, & K. Rice (Eds.), *The Blackwell companion to phonology* (Vol. 3, pp. 2258-2282). Malden, MA: Wiley Blackwell.
- Kashino, M., Wieringen, A. v., & Pols, L. C. (1992). *Cross-languages differences in the identification of intervocalic stop consonants by Japanese and Dutch listeners*. Paper presented at the Second International Conference on Spoken Language Processing.
- Katayama, M. (1998). *Optimality Theory and Japanese loanword phonology*. Doctoral dissertation, University of California, Santa Cruz,
- Kaufman, S. B., DeYoung, C. G., Gray, J. R., Jiménez, L., Brown, J., & Mackintosh, N. (2010). Implicit learning as an ability. *Cognition*, 116(3), 321-340.
- Kawagoe, I. (2015). The phonology of sokuon, or geminate obstruents. In H. Kubozono (Ed.), *Handbook of Japanese phonetics and phonology* (pp. 79–119). Berlin: De Gruyter Mouton.
- Kawagoe, I., & Arai, M. (2002). Syakuyoogo ni okeru sokuon [consonant gemination in loanwords]. *Onsei Kenkyu*, 6(1), 53-66.
- Kawahara, S. (2006). A Faithfulness Ranking Projected from a Perceptibility Scale: The Case of [+ Voice] in Japanese. *Language*, 82(3), 536-574.
- Kawahara, S. (2010). Modes of phonological judgment. *Manuscript. Rutgers University*. (Talk presented at the University of Pennsylvania).
- Kawahara, S. (2011). Japanese loanword devoicing revisited: a rating study. *Natural Language & Linguistic Theory*, 29(3), 705-723. doi:10.1007/s11049-011-9131-7

- Kawahara, S. (2015). The phonetics of sokuon, obstruent geminates. In H. Kubozono (Ed.), *Handbook of Japanese phonetics and phonology* (pp. 43-73). Berlin: De Gruyter Mouton.
- Kawahara, S. (2016). Japanese has syllables: a reply to Labrune. *Phonology*, 33(1), 169-194.
- Kay, G. (1995). English loanwords in Japanese. *World Englishes*, 14(1), 67-76.
- Kearns, K. (1990). A note on the glottal fricative in Maori. *Te Reo*, 33, 65-81.
- Kenstowicz, M. (2007). Salience and similarity in loanword adaptation: a case study from Fijian. *Language Sciences*, 29(2-3), 316-340.
- Kenstowicz, M., & Suchato, A. (2006). Issues in loanword adaptation: A case study from Thai. *Lingua*, 116(7), 921-949.
- Kilpatrick, A. J., Kawahara, S., Bundgaard-Nielsen, R. L., Baker, B. J., & Fletcher, J. (2018). *Japanese coda [m] elicits both perceptual assimilation and epenthesis*. Paper presented at the Proc. ISAPh 2018 International Symposium on Applied Phonetics.
- Kindaichi, H., & Akinaga, K. (2014). *Sinmeikai Japanese accent dictionary* (2nd ed.). Tokyo: Sanseido.
- Kizach, J. (2014). *Analyzing Likert-scale data with mixed-effects linear models: a simulation study*. Paper presented at the Linguistic Evidence, Tübingen.
- Kojic-Sabo, I., & Lightbown, P. M. (1999). Students' approaches to vocabulary learning and their relationship to success. *The Modern Language Journal*, 83(2), 176-192.
- Koo, J. H., & Homma, Y. (1989). Consonant Gemination in English Loanwords of Japanese. *The Journal of the Association of Teachers of Japanese*, 23(2), 125-134.
- Krashen, S. D. (1981). *Second language acquisition and second language learning*. Oxford: Pergamon.
- Kubozono, H. (2001). Epenthetic vowels and accent in Japanese: Facts and paradoxes. In J. van de Weijer & T. Nishihara (Eds.), *Issues in Japanese phonology and morphology* (pp. 113-142). Berlin, New York: De Gruyter Mouton.
- Kubozono, H. (2015). Loanword phonology. In H. Kubozono (Ed.), *Handbook of Japanese phonetics and phonology* (pp. 313-361). Berlin: De Gruyter Mouton.
- Kubozono, H., Itô, J., & Mester, A. (2008). *Consonant gemination in Japanese loanword phonology*. Paper presented at the 18th International Congress of Linguistics, Seoul.
- Kuhl, P. K. (1991). Human adults and human infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not. *Perception & Psychophysics*, 50(2), 93-107.

- Kuhl, P. K. (1993). Innate predispositions and the effects of experience in speech perception: The native language magnet theory. In B. de Boysson-Bardies, S. de Schonen, P. Jusczyk, P. MacNeilage, & J. Morton (Eds.), *Developmental neurocognition: Speech and face processing in the first year of life* (pp. 259-274). Hingham, MA: Kluwer Academic Press.
- Kuhl, P. K., Conboy, B. T., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M., & Nelson, T. (2008). Phonetic learning as a pathway to language: new data and native language magnet theory expanded (NLM-e). *Philosophical Transactions of the Royal Society B: Biological Sciences*, 363(1493), 979-1000.
- Kuhl, P. K., Conboy, B. T., Padden, D., Nelson, T., & Pruitt, J. (2005). Early speech perception and later language development: implications for the “Critical Period”. *Language Learning and Development*, 1(3-4), 237-264.
- Kuperman, V., Bertram, R., & Baayen, R. H. (2010). Processing trade-offs in the reading of Dutch derived words. *Journal of Memory and Language*, 62(2), 83-97.
- Lagrou, E., Hartsuiker, R., & Duyck, W. (2013). The influence of sentence context and accented speech on lexical access in second-language auditory word recognition. *Bilingualism: Language and Cognition*, 16(3), 508-517.
- Lentz, T. O., & Kager, R. W. (2015). Categorical phonotactic knowledge filters second language input, but probabilistic phonotactic knowledge can still be acquired. *Language and Speech*, 58(3), 387-413.
- Levy, B. J., McVeigh, N. D., Marful, A., & Anderson, M. C. (2007). Inhibiting your native language: The role of retrieval-induced forgetting during second-language acquisition. *Psychological Science*, 18(1), 29-34. doi:10.1111/j.1467-9280.2007.01844.x
- Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R news*, 2(3), 18-22.
- Likert, R. (1932). A technique for the measurement of attitudes. *Archives of psychology*, 140, 1-55.
- Linck, J. A., Kroll, J. F., & Sunderman, G. (2009). Losing access to the native language while immersed in a second language: Evidence for the role of inhibition in second-language learning. *Psychological Science*, 20(12), 1507-1515. doi:doi.org/10.1111/j.1467-9280.2009.02480.x
- Lovins, J. B. (1975). *Loanwords and the phonological structure of Japanese*. Bloomington: Indiana University Linguistic Club.

- Lüdtke, D. (2018). ggeffects: Create tidy data frames of marginal effects for 'ggplot' from model outputs. *R package version 0.3*.
- Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.
- MacKain, K. S., Best, C. T., & Strange, W. (1981). Categorical perception of English /r/ and /l/ by Japanese bilinguals. *Applied Psycholinguistics*, 2(4), 369-390.
- Maekawa, K., Koiso, H., Furui, S., & Isahara, H. (2000). *Spontaneous Speech Corpus of Japanese*. Paper presented at the LREC.
- Major, R. C. (1992). Losing English as a first language. *The Modern Language Journal*, 76(2), 190-208.
- Marcus, G. F., Vijayan, S., Rao, S. B., & Vishton, P. M. (1999). Rule learning by seven-month-old infants. *Science*, 283(5398), 77-80.
- Marian, V., Blumenfeld, H. K., & Boukrina, O. V. (2008). Sensitivity to phonological similarity within and across languages. *Journal of psycholinguistic research*, 37(3), 141-170.
- Marian, V., & Spivey, M. (2003a). Bilingual and monolingual processing of competing lexical items. *Applied Psycholinguistics*, 24(2), 173.
- Marian, V., & Spivey, M. (2003b). Competing activation in bilingual language processing: Within-and between-language competition. *Bilingualism*, 6(2), 97.
- Martin, C. S., Mullennix, J. W., Pisoni, D. B., & Summers, W. V. (1989). Effects of talker variability on recall of spoken word lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15(4), 676-684.
- Massaro, D. W. (1974). Perceptual units in speech recognition. *Journal of Experimental Psychology* 102(2), 199-208.
- Massaro, D. W., & Cohen, M. M. (1983). Phonological context in speech perception. *Perception & Psychophysics* 34(4), 338-348.
- Mattys, S. L., & Jusczyk, P. W. (2001). Do infants segment words or recurring contiguous patterns? *Journal of Experimental Psychology: Human Perception and Performance*, 27(3), 644.
- Mattys, S. L., Jusczyk, P. W., Luce, P. A., & Morgan, J. L. (1999). Phonotactic and prosodic effects on word segmentation in infants. *Cognitive Psychology*, 38(4), 465-494.
- Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, 82(3), B101-B111.

- McCandliss, B. D., Fiez, J. A., Protopapas, A., Conway, M., & McClelland, J. L. (2002). Success and failure in teaching the [r]-[l] contrast to Japanese adults: Tests of a Hebbian model of plasticity and stabilization in spoken language perception. *Cognitive, Affective, & Behavioral Neuroscience*, 2(2), 89-108.
- McGill, R., Tukey, J. W., & Larsen, W. A. (1978). Variations of box plots. *The American Statistician*, 32(1), 12-16.
- McQueen, J. M. (1998). Segmentation of continuous speech using phonotactics. *Journal of Memory and Language*, 39(1), 21-46.
- Metsala, J. L. (1999). Young children's phonological awareness and nonword repetition as a function of vocabulary development. *Journal of Educational Psychology*, 91(1), 3.
- Miao, R. (2005). *Loanword adaptation in Mandarin Chinese: Perceptual, phonological and sociolinguistic factors*. Doctoral dissertation, Stony Brook University.
- Mintz, T. H. (2002). Category induction from distributional cues in an artificial language. *Memory & Cognition*, 30(5), 678-686.
- Mintz, T. H., Wang, F. H., & Li, J. (2014). Word categorization from distributional information: Frames confer more than the sum of their (Bigram) parts. *Cognitive psychology*, 75, 1-27.
- Misyak, J. B., & Christiansen, M. H. (2007). *Extending statistical learning farther and further: Long-distance dependencies, and individual differences in statistical learning and language*. Paper presented at the Proceedings of the Annual Meeting of the Cognitive Science Society.
- Misyak, J. B., & Christiansen, M. H. (2012). Statistical learning and language: An individual differences study. *Language Learning*, 62(1), 302-331.
- Mitchel, A. D., & Weiss, D. J. (2010). What's in a face? Visual contributions to speech segmentation. *Language and Cognitive Processes*, 25(4), 456-482.
- Mitchell, R., Myles, F., & Marsden, E. (2019). *Second language learning theories: Fourth edition*. Retrieved from ProQuest Ebook Central <https://ebookcentral.proquest.com>
- Monahan, P. J., Takahashi, E., Nakao, C., & Idsardi, W. J. (2009). Not all epenthetic contexts are equal: Differential effects in Japanese illusory vowel perception. *Japanese/Korean Linguistics*, 17, 391-405.
- Mora, J. C., & Nadeu, M. (2012). L2 effects on the perception and production of a native vowel contrast in early bilinguals. *International Journal of Bilingualism*, 16(4), 484-500.



- Moreton, E. (2002). *Phonological grammar in speech perception*. Doctoral dissertation, University of Massachusetts Amherst.
- Moreton, E., & Amano, S. (1999). *Phonotactics in the perception of Japanese vowel length: Evidence for long-distance dependencies*. Paper presented at the Sixth European Conference on Speech Communication and Technology.
- Morita, T. (2018). *Unsupervised learning of lexical subclasses from phonotactics*. Doctoral dissertation, Massachusetts Institute of Technology.
- Mullennix, J. W., Pisoni, D. B., & Martin, C. S. (1989). Some effects of talker variability on spoken word recognition. *Journal of the Acoustical Society of America*, 85(1), 365-378.
- Munson, B., Edwards, J., & Beckman, M. E. (2005). Phonological knowledge in typical and atypical speech–sound development. *Topics in language disorders*, 25(3), 190-206.
- Nasu, A. (2015). The phonological lexicon and mimetic phonology. In H. Kubozono (Ed.), *Handbook of Japanese phonetics and phonology* (pp. 253-288). Berlin: De Gruyter Mouton.
- National Institute for Japanese Language and Linguistics. (2011). *The Balanced Corpus of Contemporary Written Japanese*.
- National Institute for Japanese Language and Linguistics. (2017). *The Corpus of Spontaneous Japanese*, 5th edition.
- Nishimura, K. (2003). *Lyman's Law in loanwords*. Master's thesis, Nagoya University.
- Nobu, T., Vardaman, J. M., & Imidas, H. b. (2006). *Imidas sekai ga wakaru jidai ga mieru: gendaijin no katakanago Ōbun ryakugo jiten = A dictionary of katakana words & abbreviations*. Tōkyō: Shūeisha.
- Nogita, A., Yamane, N., & Bird, S. (2013). The Japanese unrounded back vowel/u/is in fact rounded central/front [u-ɣ]. *Ultrafest VI Program and Abstract Booklet*, 39-42.
- Norman, G. (2010). Likert scales, levels of measurement and the “laws” of statistics. *Adv Health Sci Educ Theory Pract*, 15(5), 625-632. doi:10.1007/s10459-010-9222-y
- Okada, S. (2008). *Gendainihongono onseigengoni okeru moorano shutsugendo 'nihongohanshi kotoba koopasu' o shiyoosita choosa kka*. Paper presented at the Phonetic society of Japan, Tokyo.
- Onishi, K. H., Chambers, K. E., & Fisher, C. (2002). Learning phonotactic constraints from brief auditory experience. *Cognition*, 83(1), B13-B23.
- Onnis, L., Waterfall, H. R., & Edelman, S. (2008). Learn locally, act globally: Learning language from variation set cues. *Cognition*, 109(3), 423-430.

- Otake, T., Yoneyama, K., Cutler, A., & Van Der Lugt, A. (1996). The representation of Japanese moraic nasals. *Journal of the Acoustical Society of America*, *100*(6), 3831-3842.
- Otaki, Y. (2012). A phonological account of vowel epenthesis in Japanese loanwords: Synchronic and diachronic perspectives. *Phonological Studies*, *15*, 35-42.
- Pacton, S., & Perruchet, P. (2008). An attention-based associative account of adjacent and nonadjacent dependency learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(1), 80-96.
- Pacton, S., Perruchet, P., Fayol, M., & Cleeremans, A. (2001). Implicit learning out of the lab: The case of orthographic regularities. *Journal of experimental psychology: General*, *130*(3), 401-426.
- Pacton, S., Sobaco, A., & Perruchet, P. (2015). Is an attention-based associative account of adjacent and nonadjacent dependency learning valid? *Acta Psychologica*, *157*, 195-199.
- Paradis, C., & LaCharité, D. (1996). Saving and cost in French loanword adaptation in Fula: Predictions of the TCRS loanword model. *McGill Working Papers in Linguistics*, *11*, 46-84.
- Pederson, E., & Guion-Anderson, S. (2010). Orienting attention during phonetic training facilitates learning. *Journal of the Acoustical Society of America*, *127*(2), EL54-EL59.
- Peperkamp, S. (2004). *A psycholinguistic theory of loanword adaptations*. Paper presented at the Annual Meeting of the Berkeley Linguistics Society.
- Pierrehumbert, J. (2001). Why phonological constraints are so coarse-grained. *Language and cognitive processes*, *16*(5-6), 691-698.
- Pintér, G. (2015). The emergence of new consonant contrasts. In H. Kubozono (Ed.), *Handbook of Japanese phonetics and phonology* (pp. 121-165). Berlin: De Gruyter Mouton.
- Pisoni, D. B. (1973). Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Perception & Psychophysics*, *13*(2), 253-260.
- Polka, L. (1991). Cross-language speech perception in adults: Phonemic, phonetic, and acoustic contributions. *Journal of the Acoustical Society of America*, *89*(6), 2961-2977. doi:10.1121/1.400734
- Polka, L. (1992). Characterizing the influence of native language experience on adult speech perception. *Perception & Psychophysics*, *52*(1), 37-52.

- Polka, L., & Werker, J. F. (1994). Developmental changes in perception of nonnative vowel contrasts. *Journal of Experimental Psychology: Human Perception and Performance*, 20(2), 421.
- Porretta, V. J., & Tucker, B. V. (2015). Perception of non-native consonant length contrast: The role of attention in phonetic processing. *Second Language Research*, 31(2), 239-265. doi:10.1177/0267658314559573
- Poser, W. J. (1984). *The phonetics and phonology of tone and intonation in Japanese*. Doctoral dissertation, Massachusetts Institute of Technology.
- Potter, C. E., Wang, T., & Saffran, J. R. (2017). Second language experience facilitates statistical learning of novel linguistic materials. *Cognitive Science*, 41(Suppl 4), 913-927. doi:10.1111/cogs.12473
- Preston, D. R., & Yamagata, A. K. (2004). Katakana representation of English loanwords- Mora conservation and variable learner strategies. *Journal of Sociolinguistics*, 8(3), 359-379.
- R Core Team. (2018). R: A language and environment for statistical computing (version 3.5.2). Retrieved from <https://www.R-project.org/>.
- Reeder, P. A., Newport, E. L., & Aslin, R. N. (2013). From shared contexts to syntactic categories: The role of distributional information in learning linguistic form-classes. *Cognitive psychology*, 66(1), 30-54.
- Reeder, P. A., Newport, E. L., & Aslin, R. N. (2017). Distributional learning of subcategories in an artificial grammar: Category generalization and subcategory restrictions. *Journal of Memory and Language*, 97, 17-29.
- Revelle, W. (2015). Package 'psych'. *The Comprehensive R Archive Network*.
- Romberg, A. R., & Saffran, J. R. (2010). Statistical learning and language acquisition. *Wiley Interdisciplinary Reviews: Cognitive Science*, 1(6), 906-914. doi:10.1002/wcs.78
- Rosen, V. M., & Engle, R. W. (1998). Working memory capacity and suppression. *Journal of Memory and Language*, 39(3), 418-436.
- Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, 274(5294), 1926-1928.
- Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996). Word segmentation: The role of distributional cues. *Journal of Memory and Language*, 35(4), 606-621.
- Saffran, J. R., & Wilson, D. P. (2003). From syllables to syntax: Multilevel statistical learning by 12-month-old infants. *Infancy*, 4(2), 273-284.

- Sagisaka, Y., & Tokuhara, Y. (1984). Phoneme duration control for speech synthesis by rule. *Trans. Inst. Electron. Commun, J67-A*, 629-636.
- Sato, N., Tajima, M., Hashimoto, M., Matsushita, T., & Sasao, Y. (2017). Development of a 50,000-word-level Japanese vocabulary size test based on written-corpus frequency data.
- Schmidt, J., Janse, E., & Scharenborg, O. (2016). Perception of Emotion in Conversational Speech by Younger and Older Listeners. *Frontiers in Psychology*, 7, 781.  
doi:10.3389/fpsyg.2016.00781
- Schouten, M. E., & van Hoesen, A. J. (1992). Modeling phoneme perception. I: Categorical perception. *Journal of the Acoustical Society of America*, 92(4, Pt 1), 1841-1855.  
doi:10.1121/1.403841
- Schulpen, B., Dijkstra, T., Schriefers, H. J., & Hasper, M. (2003). Recognition of interlingual homophones in bilingual auditory word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, 29(6), 1155.
- Shahin, A. J., Bishop, C. W., & Miller, L. M. (2009). Neural mechanisms for illusory filling-in of degraded speech. *Neuroimage*, 44(3), 1133-1143.  
doi:10.1016/j.neuroimage.2008.09.045
- Shea, C. E., & Curtin, S. (2011). Experience, representations and the production of second language allophones. *Second Language Research*, 27(2), 229-250.
- Sheldon, A., & Strange, W. (1982). The acquisition of /r/ and /l/ by Japanese learners of English: Evidence that speech production can precede speech perception. *Applied Psycholinguistics*, 3(3), 243-261.
- Shibatani, M. (1990). *The languages of Japan*. Cambridge: Cambridge University Press.
- Shinohara, S. (1997). *Analyse phonologique de l'adaptation japonaise de mots étrangers*. Doctoral dissertation, ANRT, Université de Lille III.
- Shirai, S. (1999). *Gemination in loans from English to Japanese*. Master's thesis, University of Washington: University of Washington.
- Shirai, S. (2012). *Gemination in loans from English to Japanese (revised Master's thesis)*. Revised Master's thesis, University of Washington.
- Shoji, S., & Shoji, K. (2014). *Vowel Epenthesis and Consonant Deletion in Japanese Loanwords from English*. Paper presented at the Proceedings of the Annual Meetings on Phonology.

- Siegelman, N., Bogaerts, L., Christiansen, M. H., & Frost, R. (2017). Towards a theory of individual differences in statistical learning. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 372(1711), 20160059.
- Siegelman, N., & Frost, R. (2015). Statistical learning as an individual ability: Theoretical perspectives and empirical evidence. *Journal of Memory and Language*, 81, 105-120.
- Smith, J. L. (2006). Loan phonology is not all perception: Evidence from Japanese loan doublets. *Japanese/Korean Linguistics*, 14, 63-74.
- Soskuthy, M., & Hay, J. (2017). Changing word usage predicts changing word durations in New Zealand English. *Cognition*, 166, 298-313. doi:10.1016/j.cognition.2017.05.032
- Steriade, D. (2001a). Directional asymmetries in place assimilation: A perceptual account. In E. Hume & K. Johnson (Eds.), *The role of speech perception in phonology* (pp. 219-250). San Diego, CA: Academic Press.
- Steriade, D. (2001b). The phonology of perceptibility effects: the P-map and its consequences for constraint organization. In K. Hanson & S. Inkelas (Eds.), *The nature of the word*. Cambridge: MIT Press.
- Steriade, D. (2008). The Phonology of Perceptibility Effects: The P-Map and Its Consequences for Constraint Organization. In K. Hanson & S. Inkelas (Eds.), *The nature of the word: Studies in honor of Paul Kiparsky* (pp. 151-179). Cambridge: MIT Press.
- Stokes, S. F., Moran, C., & George, A. (2013). Nonword repetition and vocabulary use in toddlers. *Topics in language disorders*, 33(3), 224-237.
- Stolcke, A. (2002). *SRILM-an extensible language modeling toolkit*. Paper presented at the Seventh international conference on spoken language processing.
- Storkel, H. L. (2001). Learning new words: Phonotactic probability in language development. *Journal of Speech, Language, and Hearing Research*, 44, 1321-1337.
- Storkel, H. L. (2003). Learning new words II: Phonotactic probability in verb learning. *Journal of Speech, Language, & Hearing Research*, 46, 1312-1323.
- Storkel, H. L., Armbrüster, J., & Hogan, T. P. (2006). Differentiating phonotactic probability and neighborhood density in adult word learning. *Journal of Speech, Language, and Hearing Research*. doi:10.1044/1092-4388(2006/085)
- Storkel, H. L., & Hoover, J. R. (2011). The influence of part-word phonotactic probability/neighborhood density on word learning by preschool children varying in expressive vocabulary. *Journal of Child Language*, 38(3), 628-643.

- Strange, W., Akahane-Yamada, R., Kubo, R., Trent, S. A., & Nishi, K. (2001). Effects of consonantal context on perceptual assimilation of American English vowels by Japanese listeners. *Journal of the Acoustical Society of America*, *109*(4), 1691-1704.
- Tabata-Sandom, M. (2015). L2 reading perceptions of learners of Japanese: The influence of the reading instruction. *Reading Matrix*, *15*, 274-289.
- Tajima, K., Kato, H., Rothwell, A., Akahane-Yamada, R., & Munhall, K. G. (2008). Training English listeners to perceive phonemic length contrasts in Japanese. *Journal of the Acoustical Society of America*, *123*(1), 397-413. doi:10.1121/1.2804942
- Takagi, N., & Mann, V. (1994). A perceptual basis for the systematic phonological correspondences between Japanese loan words and their English source words. *Journal of Phonetics*, *22*(4), 343-356.
- Takagi, N., & Mann, V. (1995). The limits of extended naturalistic exposure on the perceptual mastery of English /r/ and /l/ by adult Japanese learners of English. *Applied Psycholinguistics*, *16*(4), 380-406.
- Takeyasu, H. (2010). Phonetic motivation for the gemination of stops in Japanese loanwords. *Kobe papers in linguistics*, *7*, 91-106.
- Tamaoka, K., & Makioka, S. (2004). Frequency of occurrence for units of phonemes, morae, and syllables appearing in a lexical corpus of a Japanese newspaper. *Behavior Research Methods, Instruments, & Computers*, *36*(3), 531-547.
- Tice, M., & Woodley, M. (2012). Pagnettes & bastries: Novice French learners show shifts in native phoneme boundaries. *UC Berkeley PhonLab Annual Report*, *8*(8).
- Toda, T. (2003). *Second language speech perception and production: Acquisition of phonological contrasts in Japanese*. Lanham, MD: University Press of America.
- Toro, J. M., Sinnett, S., & Soto-Faraco, S. (2005). Speech segmentation by statistical learning depends on attention. *Cognition*, *97*(2), B25-B34.
- Toro, J. M., Sinnett, S., & Soto-Faraco, S. (2011). Generalizing linguistic structures under high attention demands. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(2), 493.
- Trapman, M., & Kager, R. (2009). The acquisition of subset and superset phonotactic knowledge in a second language. *Language Acquisition*, *16*(3), 178-221.
- Tsao, F. M., Liu, H. M., & Kuhl, P. K. (2004). Speech perception in infancy predicts language development in the second year of life: A longitudinal study. *Child Development*, *75*(4), 1067-1084.
- Tsujimura, N. (1996). *An Introduction to Japanese linguistics*. Cambridge, MA: Blackwell.

- Uffmann, C. (2006). Epenthetic vowel quality in loanwords: Empirical and formal issues. *Lingua*, 116(7), 1079-1111.
- van Hessen, A. J., & Schouten, M. E. (1992). Modeling phoneme perception. II: A model of stop consonant discrimination. *Journal of the Acoustical Society of America*, 92(4 Pt 1), 1856-1868. doi:10.1121/1.403842
- Vance, T. J. (1987). *An introduction to Japanese phonology*. Albany: UNY Press.
- Vance, T. J. (2008). *The sounds of Japanese*. Cambridge, U.K: Cambridge University Press.
- Vitevitch, M. S., & Luce, P. A. (1998). When words compete: Levels of processing in perception of spoken words. *Psychological Science*, 9(4), 325-329.
- Vitevitch, M. S., & Luce, P. A. (1999). Probabilistic phonotactics and neighborhood activation in spoken word recognition. *Journal of Memory and Language*, 40(3), 374-408.
- Vitevitch, M. S., & Luce, P. A. (2005). Increases in phonotactic probability facilitate spoken nonword repetition. *Journal of Memory and Language*, 52(2), 193-204.
- Vitevitch, M. S., Luce, P. A., Charles-Luce, J., & Kemmerer, D. (1997). Phonotactics and syllable stress: Implications for the processing of spoken nonsense words. *Language and Speech*, 40(1), 47-62.
- Wang, X. (2017). *Exploring the role of attitudes in new dialect formation in Hohhot, China* Doctoral dissertation, University of Canterbury.
- Warren, R. M. (1970). Perceptual restoration of missing speech sounds. *Science*, 167(3917), 392-393.
- Webb, S. (2008). Receptive and productive vocabulary sizes of L2 learners. *Studies in Second Language Acquisition*, 79-95.
- Weber, A., & Cutler, A. (2004). Lexical competition in non-native spoken-word recognition. *Journal of Memory and Language*, 50(1), 1-25.
- Weber, A., & Cutler, A. (2006). First-language phonotactics in second-language listening. *Journal of the Acoustical Society of America*, 119(1), 597-607.
- Weinberger, S. (1994). Functional and phonetic constraints on second language phonology. In Yavas. M (Ed.), *First and second language phonology* (pp. 283-302). San Diego, CA: Singular Publishing Group.
- Weiss, D. J., Gerfen, C., & Mitchel, A. D. (2009). Speech segmentation in a simulated bilingual environment: A challenge for statistical learning? *Language Learning and Development*, 5(1), 30-49.

- Werker, J. F., & Logan, J. S. (1985). Cross-language evidence for three factors in speech perception. *Perception & Psychophysics*, 37(1), 35-44.
- Werker, J. F., & Tees, R. C. (1984). Phonemic and phonetic factors in adult cross-language speech perception. *Journal of the Acoustical Society of America*, 75(6), 1866-1878. doi:10.1121/1.390988
- Wilson, C., & Davidson, L. (2013). *Bayesian analysis of non-native cluster production*. Paper presented at the NELS Cambridge.
- Wright, M. N., & Ziegler, A. (2015). ranger: a fast implementation of random forests for high dimensional data in C++ and R (arXiv:1508.04409).
- Yamada, R. A., & Tohkura, Y. I. (1992). The effects of experimental variables on the perception of American English /r/ and /l/ by Japanese listeners. *Perception & Psychophysics*, 52(4), 376-392.
- Yip, M. C. (2015). Spoken word recognition of Chinese words in continuous speech. *Journal of Psycholinguistic Research*, 44(6), 775-787.
- Yoshida, Y. (2006). Accents in Tokyo and Kyoto Japanese vowel quality in terms of duration and licensing potency. *SOAS (School of Oriental and African Studies) Working Papers in Linguistics*, 14, 249-264.
- Zamuner, T. S., Gerken, L., & Hammond, M. (2000). Akamatsu, T. (2000). *Japanese phonology: A functional approach* (Vol. 38). Munich, Germany: Lincom Europa.



## Appendix A: List of items for pilot study

Shading marks the pronunciation type that participants are expected to give high ratings for the structure.

<b>Voiceless Stimuli</b>		Original Structure	Pronunciation Types: Representation of the target word					
Word Structure		CVC	CVCCV	CVCV	CVVCV	CVCCV		CVC
		1 syllable	3 morae	2 morae	3 morae	3 morae		1 syllable
		Target Word (Source Word)	conforming to loanword phonology	no gemination	vowel lengthening	inappropriate epenthetic	unexpected vowel	default vowel [u]
Phonological Regulation in Loanword	Acceptability	N/A	✓	✗	✗	✗	✗	✗
	Gemination	N/A	✓	✗	✗	✓	✓	✗
	Quality of Epenthetic Vowel	N/A	✓	✓	✓	✗	✗	✗
	Phonologically Legal	N/A	✓	✓	✓	✓	✗	✗
<b>Lax Vowel</b>	<b>Key Word</b>	<b>/p/-Set</b>						
[ɪ]	KIT	pip[pɪp]	/pippu/	/pipu/	/piipu/	/pippo/	N/A	/pip/
		nip[nɪp]	/nippu/	/nipu/	/niipu/	/nippo/	N/A	/nip/
[e]	DRESS	kip[kɪp]	/kippu/	/kipu/	/kiipu/	/kippo/	N/A	/kip/
[æ]	TRAP	pep[pep]	/peppu/	/pepu/	/peepu/	/peppo/	N/A	/pep/
		chap[tʃæp]	/tʃappu/	/tʃapu/	/tʃaapu/	/tʃappo/	N/A	/tʃap/
		dap[dæp]	/dappu/	/dapu/	/daapu/	/dappo/	N/A	/dap/
		nap[næp]	/nappu/	/napu/	/naapu/	/nappo/	N/A	/nap/
[ʌ]	STRUT	sup[sʌp]	/sappu/	/sapu/	/saapu/	/sappo/	N/A	/sap/
		pup[pʌp]	/pappu/	/papu/	/paapu/	/pappo/	N/A	/pap/
[ɒ]	LOT	lop[lɒp]	/roppu/	/ropu/	/roopu/	/roppo/	N/A	/rop/
<b>/t/-Set</b>								
[ɪ]	KIT	lit[lɪt]	/rittō/	/rito/	/riito/	N/A	/rittu/	/rit/
		zit[zɪt]	/zitto/	/zito/	/ziito/	N/A	/zittu/	/zit/
[e]	DRESS	bet[bet]	/betto/	/beto/	/beeto/	N/A	/bettu/	/bet/
[æ]	TRAP	dat[dæt]	/datto/	/dato/	/daato/	N/A	/dattu/	/dat/
		tat[tæt]	/tatto/	/tato/	/taato/	N/A	/tattu/	/tat/
[ʌ]	STRUT	jut[dʒʌt]	/dʒatto/	/dʒato/	/dʒaato/	N/A	/dʒattu/	/dʒat/
		gut[gʌt]	/gatto/	/gato/	/gaato/	N/A	/gattu/	/gat/
		jot[dʒɒt]	/dʒotto/	/dʒoto/	/dʒooto/	N/A	/dʒottu/	/dʒɒt/
[ɒ]	LOT	sot[sɒt]	/sotto/	/soto/	/sooto/	N/A	/sottu/	/sot/
		tot[tɒt]	/totto/	/toto/	/tooto/	N/A	/tottu/	/tot/
<b>/k/-Set</b>								
[ɪ]	KIT	lick[lɪk]	/rikku/	/riku/	/riiku/	/rikko/	N/A	/rik/
		nick[nɪk]	/nikku/	/niku/	/niiku/	/nikko/	N/A	/nik/
[e]	DRESS	peck[pek]	/pekku/	/peku/	/peeku/	/pekko/	N/A	/pek/
		beck[bek]	/bekku/	/beku/	/beeku/	/bekko/	N/A	/bek/
[æ]	TRAP	hack[hæk]	/hakku/	/haku/	/haaku/	/hakko/	N/A	/hak/
		tack[tæk]	/takku/	/taku/	/taaku/	/takko/	N/A	/tak/
[ʌ]	STRUT	puck[pʌk]	/pakku/	/paku/	/paaku/	/pakko/	N/A	/pak/
[ɒ]	LOT	hock[hɒk]	/hokku/	/hoku/	/hooku/	/hokko/	N/A	/hok/
		chock[tʃɒk]	/tʃokku/	/tʃoku/	/tʃooku/	/tʃokko/	N/A	/tʃok/
		sock[sɒk]	/sokku/	/soku/	/sooku/	/sokko/	N/A	/sok/

<b>Voiced Stimuli</b>		Original Structure	Pronunciation Types: Representation of the target word					
			CVC 1 syllable	CVCCV 3 morae	CVCV 2 morae	CVVCV 3 morae	CVCV 2 morae	CVC 1 syllable
Word Structure		Target Word (Source Word)	gemination	conforming to loanword phonology	vowel lengthening	inappropriate		only vowel phoneme substitution
						unexpected vowel	default vowel [u]	
Phonological Regulation in Loanword	Acceptability	N/A	✓	✓	✗	✗	✗	✗
	Voi Gemination	N/A	✗	✓	✓	✓	✓	✗
	Quality of Epenthetic Vowel	N/A	✓	✓	✓	✗	✗	✗
	Phonologically Legal	N/A	✓	✓	✓	✓	✗	✗
Lax Vowel	Key Word							
<b>/b/-Set</b>								
[ɪ]	KIT	jib[dʒɪb]	/dʒɪbbu/	/dʒɪbu/	/dʒɪbu/	/dʒɪbo/	N/A	/dʒɪb/
		nib[nɪb]	/nɪbbu/	/nɪbu/	/nɪbu/	/nɪbo/	N/A	/nɪb/
[æ]	TRAP	nab[næb]	/nabbu/	/nabu/	/naabu/	/nabo/	N/A	/tab/
		gab[ɡæb]	/ɡabbu/	/ɡabu/	/ɡaabu/	/ɡabo/	N/A	/gab/
[ʌ]	STRUT	dub[dʌb]	/dabbu/	/dabu/	/daabu/	/dabo/	N/A	/dab/
		bub[bʌb]	/babbu/	/babu/	/baabu/	/babo/	N/A	/bab/
[ɒ]	LOT	cob[kɒb]	/kobbu/	/kobu/	/koobu/	/kobo/	N/A	/kob/
		hob[hɒb]	/hobbu/	/hobu/	/hoobu/	/hobo/	N/A	/hob/
		gob[gɒb]	/gobbu/	/gobu/	/goobu/	/gobo/	N/A	/gob/
		mob[mɒb]	/mobbu/	/mobu/	/moobu/	/mobo/	N/A	/mob/
<b>/d/-Set</b>								
[ɪ]	KIT	hid[hɪd]	/hiddo/	/hido/	/hiido/	N/A	/hidu/	/hid/
[e]	DRESS	med [med]	/meddo/	/medo/	/meedo/	N/A	/medu/	/med/
		zed [zed]	/zeddo/	/zedo/	/zeedo/	N/A	/zedu/	/zed/
[æ]	TRAP	tad[tæd]	/taddo/	/tado/	/taado/	N/A	/tadu/	/tad/
[ʌ]	STRUT	cud[kʌd]	/kaddo/	/kado/	/kaado/	N/A	/kadu/	/kad/
		mud[mʌd]	/maddo/	/mado/	/maado/	N/A	/madu/	/mad/
[ɒ]	LOT	hod[hɒd]	/hoddo/	/hodo/	/hoodo/	N/A	/hodu/	/hod/
		cod[kɒd]	/koddoo/	/kodo/	/koodoo/	N/A	/kodu/	/kod/
		nod[nɒd]	/noddo/	/nodo/	/noodoo/	N/A	/nodu/	/nod/
		tod[tɒd]	/toddo/	/todo/	/toodo/	N/A	/todu/	/tod/
<b>/g/-Set</b>								
[e]	DRESS	keg[keg]	/keggu/	/kegu/	/keegu/	/kego/	N/A	/keg/
		meg[meg]	/meggu/	/megu/	/meegu/	/mego/	N/A	/meg/
		neg[neg]	/neggu/	/negu/	/neegu/	/nego/	N/A	/neg/
[æ]	TRAP	nag[næg]	/naggu/	/nagu/	/naagu/	/nago/	N/A	/nag/
		sag[sæg]	/saggu/	/sagu/	/saagu/	/sago/	N/A	/sag/
[ʌ]	STRUT	dug[dʌg]	/daggu/	/dagu/	/daagu/	/dago/	N/A	/dag/
[ɒ]	LOT	bog[bɒg]	/boggu/	/bogu/	/boogu/	/bogo/	N/A	/bog/
		cog[kɒg]	/koggu/	/kogu/	/koogu/	/kogo/	N/A	/kog/
		hog[hɒg]	/hoggu/	/hogu/	/hoogu/	/hogo/	N/A	/hog/
		tog[tɒg]	/toggu/	/togu/	/toogu/	/togo/	N/A	/tog/

## Appendix B: List of items for well-formedness judgment task

<b>Voiceless Stimuli</b>		Original Structure	Pronunciation Types: Representation of the target word				
			CVC	CVCCV	CVCV	CVCCV	CVC
Word Structure		1 syllable	3 morae	2 morae	3 morae		1 syllable
		Target Word (Source Word)	conforming to loanword phonology	no gemination	inappropriate epenthetic		only vowel phoneme
					unexpected vowel	default vowel [u]	
Phonological Regulation in Loanword	Acceptability	N/A	✓	✗	✗	✗	✗
	Gemination	N/A	✓	✗	✓	✓	✗
	Quality of Epenthetic Vowel	N/A	✓	✓	✗	✗	✗
	Phonologically Legal	N/A	✓	✓	✓	✗	✗
<b>Lax Vowel</b>		<b>Key Word</b>		<b>/p/-Set</b>			
[ɪ]	KIT	pip[ɪp]	/pippu/	/pipu/	/pippo/	N/A	/pip/
		nip[nɪp]	/nippu/	/nipu/	/nippo/	N/A	/nip/
[e]	DRESS	kip[kɪp]	/kippu/	/kipu/	/kippo/	N/A	/kip/
[æ]	TRAP	pep[ɪp]	/peppu/	/pepu/	/peppo/	N/A	/pep/
		chaf[ɪæp]	/ʃappu/	/ʃapu/	/ʃappo/	N/A	/ʃap/
		dap[dæp]	/dappu/	/dapu/	/dappo/	N/A	/dap/
		nap[næp]	/nappu/	/napu/	/nappo/	N/A	/nap/
[ʌ]	STRUT	sup[sʌp]	/sappu/	/sapu/	/sappo/	N/A	/sap/
		pup[pʌp]	/pappu/	/papu/	/pappo/	N/A	/pap/
[ɒ]	LOT	lop[lɒp]	/roppu/	/ropu/	/roppo/	N/A	/rop/
<b>/t/-Set</b>							
[ɪ]	KIT	lit[lɪt]	/ritto/	/rito/	N/A	/rittu/	/rit/
		zit[zɪt]	/zitto/	/zito/	N/A	/zittu/	/zit/
[e]	DRESS	bet[bet]	/betto/	/beto/	N/A	/bettu/	/bet/
[æ]	TRAP	dat[dæt]	/datto/	/dato/	N/A	/dattu/	/dat/
		tat[tæt]	/tatto/	/tato/	N/A	/tattu/	/tat/
[ʌ]	STRUT	jut[dʒʌt]	/dʒatto/	/dʒato/	N/A	/dʒattu/	/dʒat/
		gut[gʌt]	/gatto/	/gato/	N/A	/gattu/	/gat/
		jot[dʒɒt]	/dʒotto/	/dʒoto/	N/A	/dʒottu/	/dʒot/
[ɒ]	LOT	sot[sɒt]	/sotto/	/soto/	N/A	/sottu/	/sot/
		tot[tɒt]	/totto/	/toto/	N/A	/tottu/	/tot/
<b>/k/-Set</b>							
[ɪ]	KIT	lick[lɪk]	/rikku/	/riku/	/rikko/	N/A	/rik/
		nick[nɪk]	/nikku/	/niku/	/nikko/	N/A	/nik/
[e]	DRESS	peck[pek]	/pekku/	/peku/	/pekko/	N/A	/pek/
		beck[bek]	/bekku/	/beku/	/bekko/	N/A	/bek/
[æ]	TRAP	hack[hæk]	/hakku/	/haku/	/hakko/	N/A	/hak/
		tack[tæk]	/takku/	/taku/	/takko/	N/A	/tak/
[ʌ]	STRUT	puck[pʌk]	/pakku/	/paku/	/pakko/	N/A	/pak/
[ɒ]	LOT	hock[hɒk]	/hokku/	/hoku/	/hokko/	N/A	/hok/
		chock[ʃɒk]	/ʃokku/	/ʃoku/	/ʃokko/	N/A	/ʃok/
		sock[sɒk]	/sokku/	/soku/	/sokko/	N/A	/sok/

<b>Voiced Stimuli</b>		Original Structure	Pronunciation Types: Representation of the target word				
			CVC	CVCCV	CVCV	CVCV	
Word Structure		1 syllable	3 morae	2 morae	2 morae		1 syllable
		Target Word (Source Word)	gemination	conforming to loanword phonology	unexpected vowel	inappropriate epenthetic default vowel [u]	only vowel phoneme substitution
Phonological Regulation in Loanword	Acceptability	N/A	✓	✓	✗	✗	✗
	Voi Gemination	N/A	✗	✓	✓	✓	✗
	Quality of Epenthetic Vowel	N/A	✓	✓	✗	✗	✗
	Phonologically Legal	N/A	✓	✓	✓	✗	✗
<b>Lax Vowel</b>	<b>Key Word</b>	<b>/b/-Set</b>					
[ɪ]	KIT	jib[dʒɪb]	/dʒɪbbu/	/dʒɪbu/	/dʒɪbo/	N/A	/dʒɪb/
		nib[nɪb]	/nɪbbu/	/nɪbu/	/nɪbo/	N/A	/nɪb/
[æ]	TRAP	nab[næb]	/nabbu/	/nabu/	/nabo/	N/A	/tab/
		gab[gaæb]	/gabbu/	/gabu/	/gabo/	N/A	/gab/
[ʌ]	STRUT	dub[dʌb]	/dabbu/	/dabu/	/dabo/	N/A	/dab/
		bub[bʌb]	/babbu/	/babu/	/babo/	N/A	/bab/
[ɒ]	LOT	cob[kɒb]	/kobbu/	/kobu/	/kobo/	N/A	/kob/
		hob[hɒb]	/hobbu/	/hobu/	/hobo/	N/A	/hob/
		gob[gɒb]	/gobbu/	/gobu/	/gobo/	N/A	/gob/
		mob[mɒb]	/mobbu/	/mobu/	/mobo/	N/A	/mob/
<b>/d/-Set</b>							
[ɪ]	KIT	hid[hɪd]	/hiddo/	/hido/	N/A	/hidu/	/hid/
[e]	DRESS	med[mɛd]	/meddo/	/medo/	N/A	/medu/	/med/
		zed[zɛd]	/zeddo/	/zedo/	N/A	/zedu/	/zed/
[æ]	TRAP	tad[tæd]	/taddo/	/tado/	N/A	/tadu/	/tad/
[ʌ]	STRUT	cud[kʌd]	/kaddo/	/kado/	N/A	/kadu/	/kad/
		mud[mʌd]	/maddo/	/mado/	N/A	/madu/	/mad/
[ɒ]	LOT	hod[hɒd]	/hoddoo/	/hodo/	N/A	/hodoo/	/hod/
		cod[kɒd]	/koddoo/	/kodo/	N/A	/kodoo/	/kod/
		nod[nɒd]	/noddoo/	/nodo/	N/A	/nodoo/	/nod/
		tod[tɒd]	/toddoo/	/todo/	N/A	/todoo/	/tod/
<b>/g/-Set</b>							
[e]	DRESS	keg[kɛg]	/keggoo/	/kegu/	/kego/	N/A	/keg/
		meg[mɛg]	/megggoo/	/megu/	/megoo/	N/A	/meg/
		neg[nɛg]	/negggoo/	/negu/	/negoo/	N/A	/neg/
[æ]	TRAP	nag[næg]	/nagggoo/	/nagu/	/nagoo/	N/A	/nag/
		sag[sæg]	/sagggoo/	/sagu/	/sagoo/	N/A	/sag/
[ʌ]	STRUT	dug[dʌg]	/duggoo/	/dagu/	/dago/	N/A	/dag/
[ɒ]	LOT	bog[bɒg]	/boggoo/	/bogu/	/bogo/	N/A	/bog/
		cog[kɒg]	/koggoo/	/kogu/	/kogo/	N/A	/kog/
		hog[hɒg]	/hoggoo/	/hogu/	/hogo/	N/A	/hog/
		tog[tɒg]	/toggoo/	/togu/	/togo/	N/A	/tog/

# Appendix C: Preregistration document for AsPredicted

**CONFIDENTIAL - FOR PEER-REVIEW ONLY**



**L2phonological\_acquisition, Christchurch NZ, 2018-2019 (#15007)**

Created: 10/11/2018 03:22 PM (PT)

Shared: 09/22/2020 03:47 PM (PT)

This pre-registration is not yet public. This anonymized copy (without author names) was created by the author(s) to use during peer-review. A non-anonymized version (containing author names) will become publicly available only if an author makes it public. Until that happens the contents of this pre-registration are confidential.

## 1) Have any data been collected for this study already?

It's complicated. We have already collected some data but explain in Question 8 why readers may consider this a valid pre-registration nevertheless.

## 2) What's the main question being asked or hypothesis being tested in this study?

The main question being asked is to what extent non-native speakers implicitly learn the complex phonological patterns of a target language through experience of using and passive exposure to that language.

We ask

- (a) What types of phonological rules are implicitly learned by L2 learners of Japanese in relation to their language ability?
- (b) Are L2 learners with high level of Japanese proficiency able to acquire fine-grained knowledge regarding the effects of voicing and place of articulation?
- (c) Are phonological constraints possible to learn from passive exposure to the target language? If any, what phonological rules are implicitly learned by monolingual -English speakers?

I will investigate their phonological adaptation of low frequency English real words into loanwords in Japanese. A set of target words using real English words are created. Each target word will have four different pronunciations. The structure of auditory stimuli has different modifications that fully or partially conform to Japanese loanword phonology.

This experiment is a word rating task, and asks participants to judge whether the pronunciation they hear is how the word would be pronounced if this was a Japanese word, by rating how confident they are on a scale of 1-5.

## 3) Describe the key dependent variable(s) specifying how they will be measured.

The dependent variable is the rating of the word heard based on a scale from 1 to 5, and z-scored within participant. This will be treated as linear unless response patterns are strongly bimodal.

## 4) How many and which conditions will participants be assigned to?

One condition: All participants will be assigned to the same condition. All participants will listen to all 240 audio stimuli.

In the experiment, each word will only be heard once by each participant. The stimuli will be presented in random order for each participant.

Words are in one of four categories: CVC, gemination, singleton, inappropriate epenthetic vowel. The word-final consonants are two types: voiced and voiceless. Each category has 60 words.

## 5) Specify exactly which analyses you will conduct to examine the main question/hypothesis.

Analyses will be mixed effects regression models.

The trial number will be z-scored.

Planned model:

response ~ pronunciation type x voicingtype x grouptype + trialnumber + (1+trialnumber+pronunciationtype+voicingtype|subject) + (1+pronunciationtype+grouptype+trialnumber |word)

pronunciation type levels: CVC / germination / singleton / wrong epenthesis

voicing type levels: voiced / voiceless

group type levels: native-Japanese / monolingual English / Japanese learners.

We will further explore the effect of place of articulation within groups that show a pronunciation type difference. This will be done by testing a POA x voicingtype interaction within the relevant group. A POA slope will also be included by subject.

POA levels: labial / alveolar / velar

If all groups show a pronunciation type difference, we will also attempt to add POA x voicingtype x group into the overall planned model.



We also have metrics relating to degree of Japanese education and exposure. These will be explored in separate models.

If the inclusion of slope leads to convergence errors, slopes that contribute least to the model will be dropped in order to obtain convergence.

We will fit all models by starting with a fully specified model (as shown above), and then removing non-significant terms one by one, using anova comparison.

**6) Describe exactly how outliers will be defined and handled, and your precise rule(s) for excluding observations.**

Participants will be excluded if their responses are atypically non-variable. Each participant's standard deviation of responses will be calculated, and participants with a standard deviation that is 2 sds below the mean will be excluded.

Participants will be excluded

- (1) if their median RT is  $< .5$  seconds.
- (2) if they indicate they have had hearing problems

**7) How many observations will be collected or what will determine sample size? No need to justify decision, but be precise about exactly how the number will be determined.**

We intend to recruit participants online with the aim of running 100 participants for the experiment.

This experiment will consist of three groups of participants: (1) 20 native speakers of Japanese in New Zealand, (2) 60 native speakers of New Zealand English taking Japanese as second language courses at universities, and (3) 20 native New Zealand English-speaking monolinguals.

**8) Anything else you would like to pre-register? (e.g., secondary analyses, variables collected for exploratory purposes, unusual analyses planned?)**

A background survey which will be included which will ask questions of the participant such as age bracket, knowledge of Japanese, and degree of exposure to Japanese. Some exploratory analysis may take place using these social variables.

We have already recruited group 1 and have started recruiting group 2 (3 people have participated by now). We will explore the data for at least with respect to questions (a) and (b). We will then conduct the experiment for group 3 to assess question (c). If the data for group 2 shows no patterns, then the data for group 3 may not be collected.

Why readers may consider this a valid pre-registration nevertheless.

The answers for pre-registration questions had created before collecting data. However, before the registration, we have recruited group 1, but the data has not analysed yet. Data collection for group 2 has just started.

## Appendix D: Questionnaires

### **Language Background Questionnaire: For English learners of Japanese**

(Note: The questionnaire was administrated by on-line web form)

**This survey is for English native speakers who are studying Japanese at universities.**

#### **A. General Information**

1. Age: Which age group do you belong to?

18 – 24     25 – 29     30 – 34     35 – 39     40+

2. Please state your gender: \_\_\_\_\_

3. Please state your nationality or community: \_\_\_\_\_

4. Is English your native language?

Yes  No

5. Please list any other languages in which you can have an everyday conversation.

( \_\_\_\_\_ )

6. Are you currently studying Japanese?

Yes  No

7. If yes, at what level? Please tick the highest level you are taking.

100 level    200 level    300 level    Honours    Master    PhD

8. Did you study Japanese in high school?

Yes  No

9. Approximately how many years have you been studying Japanese?

Less than a year    1 year    2 years    3years    More than 3years

10. Have you ever been to Japan? If your answer is 'no', please go to question 15.

Yes  No

11. If yes, how many times have you been to Japan?

Once     Twice    Three times    Four times    More than four times

12. If yes, how long was your visit to Japan?

(If you visited to Japan more than one time, how long was your longest stay?)

Less than 1 month     Less than 6 months    1year    2 years    More than 2 years

13. If you have been to Japan, what was the purpose of the visit?

(If more than one time, please check each box that applies)

Sightseeing    Business    Home stay    Visiting relatives

14. If you have been to Japan, when did you last visit Japan?

Less than 1 month ago  More than 1 month but less than 3 months ago  More than 3 months but less than 6 months ago  More than 6 months but less than 1 year ago  More than 1 year ago

15. Are you currently studying Linguistics? Or have you studied Linguistics in the past?

Yes  No

16. If yes, at what level? Please tick the highest level.

100 level  200 level  300 level  Honours  Master  PhD

17. Do you have a history of any speech, language or hearing impairment that you are aware of?

Yes  No

### **B. Degree of knowledge of Japanese**

18. How well can you speak Japanese?

Very well (I can talk about almost anything in Japanese)

Well (I can talk about many things in Japanese)

Fairly well (I can talk about some things in Japanese)

Not very well (I can only talk about simple/basic things in Japanese)

No more than a few words or phrases

Not at all

19. How well can you understand Japanese?

Very well (I can understand almost anything said in Japanese)

Well (I can understand many things said in Japanese)

Fairly well (I can understand some things said in Japanese)

Not very well (I can only understand simple/basic things in Japanese)

No more than a few words or phrases

Not at all

20. How well can you read Japanese?

I can read my course book without a dictionary.

I can read my course book with a dictionary.

I can /read Hiragana, Katakana and a bit of Kanji.

I can read Hiragana and Katakana.

I can read only Hiragana.

Not at all



4. Do you know any “rules” about how to pronounce English words that have been borrowed into Japanese? Please explain them.

( )

### **C. Degree of Exposure to Japanese**

#### **C1. How often do you do the following?**

1. Read/browse Japanese written/drawn cartoons or magazines

Never  Once a semester  Monthly  Weekly  Daily

2. Watch Japanese TV programs or Japanese movies on the internet (with English subtitles)

Never  Once a semester  Monthly  Weekly  Daily

3. Play Japanese video games (e.g., Pokémon, Mario brothers)

Never  Once a semester  Monthly  Weekly  Daily

4. Access websites that contain Japanese language resources

Never  Once a semester  Monthly  Weekly  Daily

5. Access websites about traditional Japanese culture, including literature and history

Never  Once a semester  Monthly  Weekly  Daily

6. Access websites about modern Japanese culture, including animation, cartoons (i.e., manga) and pop-culture.

Never  Once a semester  Monthly  Weekly  Daily

7. Attend Japanese social events (e.g., language exchange group, a social activity)

Never  Once a semester  Monthly  Weekly  Daily

8. Interact with Japanese people (e.g., regular social interactions)

Never  Once a semester  Monthly  Weekly  Daily

9. Go to Japanese restaurants

Never  Once a semester  Monthly  Weekly  Daily

10. Buy Japanese take away

Never  Once a semester  Monthly  Weekly  Daily

11. Work at a Japanese restaurant or grocery store

Never  Once a semester  Monthly  Weekly  Daily

#### **C2. Who speaks Japanese in your home or family at the present time?**

1. Grandmother  Yes  No

- 2. Grandfather     Yes     No
- 3. Mother         Yes     No
- 4. Father         Yes     No
- 5. Brother, Sister     Yes  No
- 6. Spouse         Yes  No
- 7. Flatmate        Yes  No
- 8. Japanese friend    Yes  No
- 9. Homestay student  Yes  No
- 10. Nobody        Yes  Other (                    )

**C3. How much do you and the other people in your home use Japanese at home?**

- Never  Very Little  Sometimes  Most of the Time  Always

**C4. How often do you hear Japanese at university at the present time?**

- Never  Once a semester  Monthly  Weekly  Daily

**C5. How often do you hear Japanese outside of university at the present time?**

- Never  Once a semester  Monthly  Weekly  Daily

**Language Background Questionnaire: For non-learners of Japanese**

**This survey is for English native speakers who are not studying Japanese at universities.**

**A. General Information**

1. Age: Which age group do you belong to?

- 18 – 24     25 – 29     30 – 34     35 – 39     +40

2. Please state your gender: \_\_\_\_\_

3. Please state your nationality or community: \_\_\_\_\_

4. Is English your native language?

- Yes  No

5. University year: which year currently do you belong to?

- First year undergraduate     Second year undergraduate     Third year undergraduate  
 Post-graduate

6. Did you study Japanese in high school? If the answer is ‘no’, please go to question 8.

- Yes  No

7. Approximately how many years have you been studying Japanese

- Less than a year     1 year     2 years     3years     More than 3years

8. Have you ever been to Japan? If the answer is ‘no’, please go to question 13.

- Yes  No

9. If yes, how many times have you been to Japan?

Once Twice Three times Four times More than four times

10. If yes, how long have you been to Japan?

(If you visited to Japan more than once, how long was your longest stay?)

Less than 1 month Less than 6months 1year 2years More than 2 years

11. If yes, what was the purpose of the visit?

(If more than once, please check each box that applies and say numbers)

Sightseeing ( ) Business ( ) Home stay ( ) Visiting relatives ( )

12. If yes, when did you last visit Japan?

Less than 1months ago 3 months ago 6 months ago 1year ago

More than 1year ago

13. Do you have a history of any speech, language or hearing impairment that you are aware of?

Yes  No

## **B. Degree of knowledge of Japanese**

1. How well can you speak Japanese?

Very well (I can talk about almost anything in Japanese)

Well (I can talk about many things in Japanese)

Fairly well (I can talk about some things in Japanese)

Not very well (I can only talk about simple/basic things in Japanese)

No more than a few words or phrases

Not at all

2. How well can you understand Japanese?

Very well (I can understand about almost anything said in Japanese)

Well (I can understand many things said in Japanese)

Fairly well (I can understand some things said in Japanese)

Not very well (I can only understand simple/basic things in Japanese)

No more than a few words or phrases

Not at all

3. Please tick all boxes that apply?

I know how to say some basic phrases (e.g., My name is ..., I am from....) in Japanese.

I know how to say some greetings in Japanese.

- I know how to say some numbers in Japanese.
- I know how to say some foods in Japanese.
- I know how to say some commands (e.g., Sit down/ Come here) in Japanese.
- I can sing a few songs in Japanese.

### **C. Degree of Exposure to Japanese**

#### **C1. How often do you do the following?**

1. Read/browse Japanese written/drawn cartoons or magazines  
 Never  Once a semester  Monthly  Weekly  Daily
2. Watch Japanese TV programs or Japanese movie on the internet (with English subtitles)  
 Never  Once a semester  Monthly  Weekly  Daily
3. Do Japanese video games (e.g., Pokémon, Mario brothers)  
 Never  Once a semester  Monthly  Weekly  Daily
4. Access websites that contain Japanese language resources  
 Never  Once a semester  Monthly  Weekly  Daily
5. Access websites about traditional Japanese culture, including literature and history  
 Never  Once a semester  Monthly  Weekly  Daily
6. Access websites about modern Japanese culture, including animation, cartoons (i.e., manga) and pop-culture.  
 Never  Once a semester  Monthly  Weekly  Daily
7. Attend to Japanese social group (e.g., language exchange group, social activity)  
 Never  Once a semester  Monthly  Weekly  Daily
8. Interact with Japanese people (e.g., regular social interactions)  
 Never  Once a semester  Monthly  Weekly  Daily
9. Go to Japanese restaurant  
 Never  Once a semester  Monthly  Weekly  Daily
10. Buy Japanese foods take away  
 Never  Once a semester  Monthly  Weekly  Daily
11. Work at Japanese restaurant or grocery store  
 Never  Once a semester  Monthly  Weekly  Daily

#### **C2. Who speaks Japanese in your home or family at the present time?**

1. Grandmother  Yes  No
2. Grandfather  Yes  No
3. Mother  Yes  No

- 4. Father  Yes  No
- 5. Brothers, Sisters  Yes  No
- 6. Spouse  Yes  No
- 7. Flatmates  Yes  No
- 8. Japanese friends  Yes  No
- 9. Homestay student  Yes  No

**C3. How much do you and your family use Japanese at home?**

- Never  Very Little  Sometimes  Most of the Time  Always

**C4. How often do you hear Japanese at university at the present time?**

- Never  Very Little  Sometimes  Most of the Time  Always

**C5. How often do you hear Japanese outside of university at the present time?**

- Never  Very Little  Sometimes  Most of the Time  Always

**Language Background Questionnaire: For native speakers of Japanese**

**A. General Information**

1. Age: Which age group do you belong to?

- 18 – 24     25 – 29     30 – 34     35 – 39     +40

2. Please state your gender: \_\_\_\_\_

3. Please state your nationality or community: \_\_\_\_\_

4. Is Japanese your native language?

- Yes  No

5. Were you born in New Zealand?

- Yes  No

6. University year: which year currently do you belong to?

- First year undergraduate  Second year undergraduate  Third year undergraduate  
 Post-graduate

7. Approximately how many years have you been in New Zealand?

- Less than a year  1 year  2 years  3 years  More than 3 years

8. Do you have a history of any speech, language or hearing impairment that you are aware of?  Yes  No

**B. Degree of knowledge of English**

1. How well can you speak English?

- Very well (I can talk about almost anything in English)  
 Well (I can talk about many things in English)

- Fairly well (I can talk about some things in English)
- Not very well (I can only talk about simple/basic things in English)
- No more than a few words or phrases
- Not at all

2. How well can you understand/read English?

- Very well (I can understand about almost anything said/written in English)
- Well (I can understand many things said/written in English)
- Fairly well (I can understand some things said/written in English)
- Not very well (I can only understand simple/basic things said/written in English)
- No more than a few words or phrases
- Not at all

# Appendix E: Preregistration document for AsPredicted

**CONFIDENTIAL - FOR PEER-REVIEW ONLY**



## Non-native sound contrasts and recognition memory for spoken words, 2019 (#27006)

Created: 08/21/2019 09:48 PM (PT)

Shared: 09/22/2020 03:44 PM (PT)

This pre-registration is not yet public. This anonymized copy (without author names) was created by the author(s) to use during peer-review. A non-anonymized version (containing author names) will become publicly available only if an author makes it public. Until that happens the contents of this pre-registration are confidential.

### 1) Have any data been collected for this study already?

No, no data have been collected for this study yet.

### 2) What's the main question being asked or hypothesis being tested in this study?

The main question being asked is whether English-speaking learners of Japanese can perceive contrasts between a single consonant (=singleton, e.g., 'kipu') and geminates (e.g., 'kippu'), and native speakers of Japanese can perceive contrasts between 'kipu' and 'kip'. Their perception might be vulnerable when task demands are high. It would be expected that increasing working memory load might slow down spoken word recognition. I also test the effect of voicing types and place of articulation in the word final position.

An experiment will be conducted. Participants will be asked to perform an auditory memory decision task on lists of stimuli that contain pairs of non-native sound contrasts. Participants hear a list of four words (e.g., 'niko', 'kippu', 'hato', 'kasa') followed by a 300ms beep. After the beep, another word (e.g., 'kipu') will be heard. They are asked to decide whether the word after the beep was in the list words before the beep by clicking 'Yes' or 'No' buttons. It is hypothesized that Japanese participants will respond poorly when stimuli involve word final contrast, whereas English-learners of Japanese will respond poorly when stimuli involve consonant length contrast. The effects of voicing types and POA on perception of non-native sound contrasts are expected.

### 3) Describe the key dependent variable(s) specifying how they will be measured.

The dependent variable is binary responses.

### 4) How many and which conditions will participants be assigned to?

One condition: All participants will be assigned to the same condition and will listen to all 240 audio-list stimuli. Each word list will only be heard once by each participant. The stimuli will be presented in random order for each participant. There are 40 target words and each target word will have three different pronunciations: CVC, geminate, singleton. The word-final consonants are four types: alveolar/velar voiced and alveolar/velar voiceless.

'Yes' stimuli contain pairs of same sound types

singleton-singleton, geminate-geminate, CVC-CVC, filler-filler

'No' stimuli contain pairs of non-native sound contrasts

singleton-geminate, geminate-singleton, singleton-CVC, CVC-singleton, singleton-filler

### 5) Specify exactly which analyses you will conduct to examine the main question/hypothesis.

Analyses will be mixed effects regression models. The trial number will be z-scored.

Planned model:

response ~ voicingtype x POA x grouptype + trialnumber + (1+trialnumber+POA +voicingtype|subject) + (1+grouptype+trialnumber |word)

POA levels: alveolar/velar

voicing type levels: voiced /voiceless

group type levels: native-Japanese/monolingual English-speakers/Japanese-learners

I also have metrics relating to degree of Japanese education and exposure. These will be explored in separate models.

If the inclusion of slope leads to convergence errors, slopes that contribute least to the model will be dropped in order to obtain convergence. I will fit all models by starting with a fully specified model, and then removing non-significant terms one by one, using anova comparison.

### 6) Describe exactly how outliers will be defined and handled, and your precise rule(s) for excluding observations.

Participants will be excluded if their responses are atypically non-variable.

(1) if their median RT is < .5 seconds.

(2) if they indicate they have had hearing problems.

### 7) How many observations will be collected or what will determine sample size? No need to justify decision, but be precise about exactly how the number will be determined.

Three groups of participants:

30 Japanese native speakers in New Zealand

30 native speakers of NZ English taking Japanese as a second language course at universities



30 native NZ English-speaking monolinguals

**8) Anything else you would like to pre-register? (e.g., secondary analyses, variables collected for exploratory purposes, unusual analyses planned?)**  
Nothing else to pre-register.



## Appendix F: Short lists for memory decision task

Note that words indicated in bold are target words, while words in non-bolded are fillers.

Alveolar Voiceless Stimuli					Alveolar Voiced Stimuli						
	Word1	Word 2	Word3	Word4	Another word		Word1	Word 2	Word3	Word4	Another word
Answer	geminate				singleton	Answer	geminate				singleton
no	<b>butto</b>	dona	nepu	kede	<b>buto</b>	no	hope	<b>buddo</b>	kima	notsu	<b>budo</b>
no	hapa	<b>detto</b>	ruku	gate	<b>deto</b>	no	baki	topa	<b>giddo</b>	nitsu	<b>gido</b>
no	bona	mihi	<b>ketto</b>	sapu	<b>keto</b>	no	<b>kuddo</b>	nozhu	pane	sota	<b>kudo</b>
no	<b>kutto</b>	nozhu	ruse	nipa	<b>kuto</b>	no	kota	<b>neddo</b>	ruchi	wamu	<b>nedo</b>
no	rani	<b>metto</b>	kuba	sepu	<b>meto</b>	no	noga	kipu	<b>muddo</b>	taze	<b>mudo</b>
no	heni	kapa	<b>mutto</b>	pogu	<b>mutu</b>	no	<b>ruddo</b>	puta	tebu	nime	<b>rudo</b>
no	<b>natto</b>	tega	pani	mofu	<b>nato</b>	no	nefu	<b>suddo</b>	kupi	bata	<b>sudo</b>
no	kema	<b>nutto</b>	gitsu	bite	<b>nuto</b>	no	peka	tozhu	<b>teddo</b>	neke	<b>tedo</b>
no	nipu	kene	<b>tetto</b>	miga	<b>teto</b>	no	<b>zaddo</b>	moki	kapu	gina	<b>zado</b>
no	<b>zatto</b>	dapu	roni	kopa	<b>zato</b>	no	puna	<b>zuddo</b>	mohe	sobu	<b>zudo</b>
	singleton				geminate		singleton				geminate
no	nime	<b>buto</b>	pogu	mota	<b>butto</b>	no	kapa	tode	<b>budo</b>	nipu	<b>buddo</b>
no	mohe	kewa	<b>deto</b>	tebu	<b>detto</b>	no	<b>gido</b>	memi	dapu	kopa	<b>giddo</b>
no	<b>keto</b>	neka	moki	sobu	<b>ketto</b>	no	gade	<b>kudo</b>	mofu	nipa	<b>kuddo</b>
no	daka	<b>kuto</b>	pane	nozhu	<b>kutto</b>	no	shoni	kema	<b>mudo</b>	gitsu	<b>muddo</b>
no	kupe	notsu	<b>meto</b>	zoka	<b>metto</b>	no	<b>nedo</b>	tega	ruku	bomi	<b>neddo</b>
no	<b>mutu</b>	nitsu	baki	zeka	<b>mutto</b>	no	meki	<b>rudo</b>	nozhu	hapa	<b>ruddo</b>
no	kuna	<b>nato</b>	sapu	taze	<b>natto</b>	no	tozhu	pene	<b>sudo</b>	miga	<b>suddo</b>
no	kipu	meda	<b>nuto</b>	ruchi	<b>nutto</b>	no	<b>tedo</b>	bona	rade	kapu	<b>teddo</b>
no	<b>teto</b>	nefu	mapa	kupi	<b>tetto</b>	no	wamu	<b>zado</b>	dona	neke	<b>zaddo</b>
no	nepu	<b>zato</b>	mona	hope	<b>zatto</b>	no	kuba	wami	<b>zudo</b>	sepu	<b>zuddo</b>
	CVC				singleton		CVC				singleton
no	moki	nepu	<b>/but/</b>	seka	<b>buto</b>	no	<b>/bud/</b>	nipu	mota	tode	<b>budo</b>
no	<b>/det/</b>	pane	ruku	mema	<b>deto</b>	no	zeka	<b>/gid/</b>	memi	nitsu	<b>gido</b>
no	sapu	<b>/ket/</b>	taze	masha	<b>keto</b>	no	pene	mapa	<b>/kud/</b>	nozhu	<b>kudo</b>
no	bima	nozhu	<b>/kut/</b>	mohe	<b>kuto</b>	no	<b>/mud/</b>	kipu	zoka	shoni	<b>mudo</b>
no	<b>/met/</b>	noka	gitsu	baki	<b>meto</b>	no	kuna	<b>/ned/</b>	gade	wamu	<b>nedo</b>
no	sepu	<b>/mut/</b>	gota	nime	<b>mutu</b>	no	meki	tebu	<b>/rud/</b>	kewa	<b>rudo</b>
no	ruchi	pogu	<b>/nat/</b>	suba	<b>nato</b>	no	<b>/sud/</b>	meda	tozhu	chine	<b>sudo</b>
no	<b>/nut/</b>	dasa	mofu	neke	<b>nuto</b>	no	nefu	<b>/ted/</b>	daka	rade	<b>tedo</b>
no	sema	<b>/tet/</b>	kupi	notsu	<b>teto</b>	no	wami	mona	<b>/zad/</b>	kapu	<b>zado</b>
no	hope	dapu	<b>/zat/</b>	tewa	<b>zato</b>	no	<b>/zud/</b>	sobu	kupe	neka	<b>zudo</b>
	singleotn				CVC		singleton				CVC
no	<b>buto</b>	nefu	kupe	tewa	<b>/but/</b>	no	seka	<b>budo</b>	wade	nozhu	<b>/bud/</b>
no	puna	<b>deto</b>	meki	tozhu	<b>/det/</b>	no	mide	sema	<b>gido</b>	nitsu	<b>/gid/</b>
no	peka	tebu	<b>keto</b>	memi	<b>/ket/</b>	no	<b>kudo</b>	notsu	teni	mema	<b>/kud/</b>
no	<b>kuto</b>	topa	rade	mofu	<b>/kut/</b>	no	dasa	<b>mudo</b>	nipu	bomi	<b>/mud/</b>
no	nozhu	<b>meto</b>	puta	gade	<b>/met/</b>	no	chine	suba	<b>nedo</b>	kapu	<b>/ned/</b>
no	pene	sota	<b>mutu</b>	gitsu	<b>/mut/</b>	no	<b>rudo</b>	gota	kipu	soze	<b>/rud/</b>
no	<b>nato</b>	kima	sapu	tode	<b>/nat/</b>	no	kima	<b>sudo</b>	taze	nepu	<b>/sud/</b>
no	kota	<b>nuto</b>	shoni	pogu	<b>/nut/</b>	no	bima	sobu	<b>tedo</b>	kide	<b>/ted/</b>
no	bata	sepu	<b>teto</b>	kene	<b>/tet/</b>	no	<b>zado</b>	wamu	noka	kede	<b>/zad/</b>
no	<b>zato</b>	dapu	gina	wami	<b>/zat/</b>	no	nipa	<b>zudo</b>	toke	ruku	<b>/zud/</b>
yes sing	nozhu	<b>buto</b>	kima	gade	<b>buto</b>	yes sing	nipu	kide	<b>budo</b>	mema	<b>budo</b>
yes sing	baki	tebu	<b>deto</b>	puta	<b>deto</b>	yes sing	<b>nedo</b>	gota	mohe	kapu	<b>nedo</b>
yes sing	<b>kuto</b>	peka	gate	sapu	<b>kuto</b>	yes sing	tewa	<b>rudo</b>	memi	sobu	<b>rudo</b>

yes sing	puna	<b>meto</b>	bite	tozhu	<b>meto</b>	yes sing	noka	mide	<b>tedo</b>	kipu	<b>tedo</b>
yes sing	ruku	hope	<b>teto</b>	gina	<b>teto</b>	yes sing	<b>zado</b>	sema	gitsu	kupe	<b>zado</b>
yes gem	<b>butto</b>	noga	kede	sepu	<b>butto</b>	yes gem	kupi	<b>buddo</b>	bima	nefu	<b>buddo</b>
yes gem	mofu	<b>detto</b>	kota	chine	<b>detto</b>	yes gem	seka	koze	<b>muddo</b>	nitsu	<b>muddo</b>
yes gem	mihi	nozu	<b>ketto</b>	bata	<b>ketto</b>	yes gem	<b>ruddo</b>	masha	bomi	dapu	<b>ruddo</b>
yes gem	<b>natto</b>	sota	kene	wamu	<b>natto</b>	yes gem	nepu	<b>teddo</b>	suba	moki	<b>teddo</b>
yes gem	topa	<b>tetto</b>	heni	pogu	<b>tetto</b>	yes gem	dasa	notsu	<b>zuddo</b>	meki	<b>zuddo</b>
yes CVC	mota	wamu	<b>ket</b>	ruchi	<b>ket</b>	yes CVC	<b>gid</b>	dona	kapu	ruse	<b>gid</b>
yes CVC	<b>met</b>	zeka	pane	sobu	<b>met</b>	yes CVC	taze	<b>kud</b>	nipa	nefu	<b>kud</b>
yes CVC	kuna	<b>mut</b>	pene	gitsu	<b>mut</b>	yes CVC	ruku	shoni	<b>mud</b>	bona	<b>mud</b>
yes CVC	toke	dapu	<b>nut</b>	mona	<b>nut</b>	yes CVC	<b>sud</b>	nepu	soze	kema	<b>sud</b>
yes CVC	<b>zat</b>	nitsu	meda	wami	<b>zat</b>	yes CVC	sapu	<b>zad</b>	tega	wami	<b>zad</b>
yes filler	pako	<b>tozhu</b>	neke	daka	<b>tozhu</b>	yes filler	nime	sepu	<b>mogo</b>	kuba	<b>mogo</b>
yes filler	mofu	wade	<b>bisho</b>	neka	<b>bisho</b>	yes filler	<b>kopa</b>	tsugo	teni	nozu	<b>kopa</b>
yes filler	<b>zoka</b>	poto	rade	wamu	<b>zoka</b>	no filler	notsu	<b>miga</b>	nemo	pani	<b>dasa</b>
no filler	beno	<b>kewa</b>	rani	nozhu	<b>mema</b>	no filler	kapa	roni	<b>zoto</b>	pogu	<b>mofu</b>
no filler	kipu	mapa	<b>nogo</b>	tode	<b>dasa</b>	no filler	<b>meyo</b>	nipu	hapa	subi	<b>dona</b>

**Velar Voiceless Stimuli**

**Velar Voiced Stimuli**

Velar Voiceless Stimuli					Velar Voiced Stimuli						
Answer	Word list				Another word	Answer	Word list				Another word
no	geminant				singleton	no	geminant			singleton	
no	kupe	nemo	<b>gekku</b>	masha	<b>geku</b>	no	<b>beggu</b>	mota	wade	napo	<b>begu</b>
no	<b>gikku</b>	noka	tsugo	wami	<b>giku</b>	no	zoka	<b>deggu</b>	bimo	toke	<b>degu</b>
no	suba	<b>hekku</b>	meyo	gade	<b>heku</b>	no	teni	daka	<b>hegggu</b>	tano	<b>hegu</b>
no	ruku	pene	<b>kekku</b>	bima	<b>keku</b>	no	<b>migggu</b>	soze	bemo	neka	<b>migu</b>
no	<b>mikku</b>	nogo	seka	rade	<b>miku</b>	no	zoto	<b>mugggu</b>	kuna	chine	<b>mugu</b>
no	sema	<b>nukku</b>	beno	poki	<b>nuku</b>	no	pako	mide	<b>noggu</b>	zeka	<b>nogu</b>
no	memi	tewa	<b>pukku</b>	nosho	<b>puku</b>	no	<b>poggu</b>	tano	bomi	meda	<b>pugu</b>
no	<b>rekku</b>	mogo	shoni	dasa	<b>reku</b>	no	kide	<b>seggu</b>	mapa	beto	<b>segu</b>
no	tode	<b>sekku</b>	mema	bisho	<b>seku</b>	no	mona	koze	<b>tegggu</b>	nibo	<b>tegu</b>
no	gota	meki	<b>zekku</b>	beno	<b>zeku</b>	no	<b>zeggu</b>	subi	kewa	nisho	<b>zegu</b>
no	singleton				geminant	no	singleton			geminant	
no	<b>geku</b>	puna	nisho	toke	<b>gekku</b>	no	meyo	<b>begu</b>	roni	seka	<b>beggu</b>
no	bata	<b>giku</b>	koze	nosho	<b>gikku</b>	no	beno	masha	<b>degu</b>	rani	<b>deggu</b>
no	mogo	chine	<b>heku</b>	kota	<b>hekku</b>	no	<b>hegu</b>	napo	mihi	suba	<b>hegggu</b>
no	<b>keku</b>	sota	bomi	napo	<b>kekku</b>	no	sema	<b>migu</b>	kene	bimo	<b>migggu</b>
no	peka	<b>miku</b>	tano	wade	<b>mikku</b>	no	pako	bima	<b>mugu</b>	gate	<b>mugggu</b>
no	gina	soze	<b>nuku</b>	bisho	<b>nukku</b>	no	<b>nogu</b>	tewa	heni	bimo	<b>noggu</b>
no	<b>puku</b>	topa	tsugo	mide	<b>pukku</b>	no	nemo	<b>pugu</b>	dasa	ruse	<b>pugggu</b>
no	beto	<b>reku</b>	kide	noga	<b>rekku</b>	no	noka	kema	<b>segu</b>	nibo	<b>seggu</b>
no	kima	bemo	<b>seku</b>	teni	<b>sekku</b>	no	<b>tegu</b>	mema	pani	zoto	<b>tegggu</b>
no	<b>zeku</b>	puta	subi	nogo	<b>zekku</b>	no	gota	<b>zegu</b>	bite	ruku	<b>zeggu</b>
no	CVC				singleton	no	CVC			singleton	
no	kuba	/gek/	toke	nogo	<b>geku</b>	no	sota	kede	/beg/	napo	<b>begu</b>
no	hapa	koze	/gik/	tsugo	<b>giku</b>	no	/deg/	rani	meyo	puna	<b>degu</b>
no	/hek/	chine	bimo	tega	<b>heku</b>	no	beto	/heg/	peka	roni	<b>hegu</b>
no	zoto	/kek/	nipa	kide	<b>keku</b>	no	bata	kene	/mig/	pako	<b>migu</b>
no	wade	mogo	/mik/	dona	<b>miku</b>	no	/mug/	pani	nemo	kota	<b>mugu</b>
no	/nuk/	soze	beno	miga	<b>nuku</b>	no	heni	/nog/	topa	bemo	<b>nogu</b>
no	nosho	/puk/	kapa	bomi	<b>puku</b>	no	bite	tano	/pug/	noga	<b>pugu</b>
no	meyo	bona	/rek/	mide	<b>reku</b>	no	/seg/	kima	ruse	nibo	<b>segu</b>
no	/sek/	kema	teni	bisho	<b>seku</b>	no	poto	/teg/	gina	mihi	<b>tegu</b>
no	kopa	/zek/	chido	subi	<b>zeku</b>	no	gate	nisho	/zeg/	puta	<b>zegu</b>

	singleton				CVC		singleton				CVC
no	rani	meyo	<b>geku</b>	zoka	/gek/	no	<b>begu</b>	pane	tano	miga	/beg/
no	<b>giku</b>	mona	kene	bisho	/gik/	no	napo	<b>deg</b>	kuba	mohe	/deg/
no	mogo	<b>heku</b>	kuna	pani	/hek/	no	neke	pako	<b>hegu</b>	bona	/heg/
no	tsugo	mapa	<b>keku</b>	ruse	/kek/	no	<b>migu</b>	dona	kupi	nisho	/mig/
no	<b>miku</b>	koze	bimo	neka	/mik/	no	ruchi	<b>mugu</b>	bemo	tega	/mug/
no-	gate	<b>nuku</b>	zoto	mota	/nuk/	no	kapa	baki	<b>nogu</b>	zoto	/nog/
no	zeka	nogo	<b>puku</b>	heni	/puk/	no	<b>pugu</b>	subi	tano	kopa	/pug/
no	<b>reku</b>	meda	bite	nosho	/rek/	no	nibo	<b>segu</b>	masha	hope	/seg/
no	tsugo	<b>seku</b>	daka	mihi	/sek/	no	beno	kema	<b>tegu</b>	moki	/teg/
no	bata	roni	<b>zeku</b>	nemo	/zek/	no	<b>zegu</b>	hapa	nime	betu	/zeg/
yes sing	<b>giku</b>	poki	zeka	bisho	<b>giku</b>	yes sing	nibo	<b>begu</b>	kema	pane	<b>begu</b>
yes sing	meyo	<b>heku</b>	daka	roni	<b>heku</b>	yes sing	kuba	nisho	<b>deg</b>	pene	<b>deg</b>
yes sing	rade	pako	<b>keku</b>	mona	<b>keku</b>	yes sing	<b>mugu</b>	teni	kapa	nosho	<b>mugu</b>
yes sing	<b>miku</b>	ruchi	bemo	kewa	<b>miku</b>	yes sing	hapa	<b>pugu</b>	bimo	tode	<b>pugu</b>
yes sing	rani	<b>seku</b>	zoka	betu	<b>seku</b>	yes sing	beno	wade	<b>zegu</b>	kopa	<b>zegu</b>
yes gem	nogo	wami	<b>gekku</b>	neka	<b>gekku</b>	yes gem	<b>beggu</b>	ruse	poto	dona	<b>beggu</b>
yes gem	<b>kekku</b>	mapa	neke	zoto	<b>kekku</b>	yes gem	nipa	<b>deggu</b>	soze	tano	<b>deggu</b>
yes gem	mota	<b>nukku</b>	nime	tsugo	<b>nukku</b>	yes gem	taze	bona	<b>muggu</b>	nemo	<b>muggu</b>
yes gem	toke	chido	<b>pukku</b>	meda	<b>pukku</b>	yes gem	<b>puggu</b>	ruko	shoni	miga	<b>puggu</b>
yes gem	<b>rekku</b>	kuna	pani	mogo	<b>rekku</b>	yes gem	napo	<b>zeggu</b>	subi	tega	<b>zeggu</b>
yes CVC	bata	<b>gik</b>	heni	pako	<b>gik</b>	yes CVC	koze	tewa	<b>heg</b>	nemo	<b>heg</b>
yes CVC	gate	bisho	<b>puk</b>	kota	<b>puk</b>	yes CVC	<b>mig</b>	kide	betu	noka	<b>mig</b>
yes CVC	<b>rek</b>	bemo	topa	gade	<b>rek</b>	yes CVC	suba	<b>nog</b>	moki	nisho	<b>nog</b>
yes CVC	bomi	<b>sek</b>	peka	nogo	<b>sek</b>	yes CVC	kupe	nosho	<b>seg</b>	mema	<b>seg</b>
yes CVC	beno	meki	<b>zek</b>	sota	<b>zek</b>	yes CVC	<b>teg</b>	dasa	bimo	kene	<b>teg</b>
yes filler	<b>sobu</b>	mihi	puna	tsugo	<b>sobu</b>	yes filler	pogu	<b>memi</b>	gota	betu	<b>memi</b>
yes filler	tebu	<b>kima</b>	baki	mogo	<b>kima</b>	yes filler	sema	nibo	<b>hope</b>	gitsu	<b>hope</b>
no filler	gina	napo	<b>sepu</b>	bite	<b>wade</b>	yes filler	<b>dapu</b>	bima	mohe	nisho	<b>dapu</b>
no filler	<b>nozu</b>	meyo	puta	kede	<b>bona</b>	no filler	kupi	<b>gitsu</b>	masha	tano	<b>ruse</b>
no filler	poto	<b>ruku</b>	noga	chine	<b>mohe</b>	no filler	seka	zoto	<b>nozhu</b>	mide	<b>puna</b>