Title
Relationship between perceptual accuracy and information measures: A cross-linguistic study

Author
Kang, Shinae

Publication Date
2015

Peer reviewed|Thesis/dissertation

Relationship between perceptual accuracy and information measures: A cross-linguistic study

by

Shinae Kang

A dissertation submitted in partial satisfaction of the
requirements for the degree of
Doctor of Philosophy

in

Linguistics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Keith A. Johnson, Chair
Professor Sharon Inkelas
Professor Susan S. Lin
Professor Robert T. Knight

Fall 2015

Relationship between perceptual accuracy and information measures: A cross-linguistic study

Abstract

Relationship between perceptual accuracy and information measures: A cross-linguistic study

by

Shinae Kang

Doctor of Philosophy in Linguistics

University of California, Berkeley

Professor Keith A. Johnson, Chair

The current dissertation studies how the information conveyed by different speech elements of English, Japanese and Korean correlates with perceptual accuracy. Two well-established information measures are used: weighted negative contextual predictability (informativity) of a speech element; and the contribution of a speech element to syllable differentiation, or functional load. This dissertation finds that the correlation between information and perceptual accuracy differs depending on both the type of information measure and the language of the listener.

To compute the information measures, Chapter 2 introduces a new corpus consisting of all the possible syllables for each of the three languages. The chapter shows that the two information measures are inversely correlated.

In Chapters 3 and 4, two perception experiments in audio-only and audiovisual modalities, respectively, are described. The experiments adopt a forced-choice-identification paradigm. In both experiments, subjects listened to VC.CV stimuli composed of spliced VC and CV chunks, and they were asked to identify the CC sequence. Multiple statistical models are constructed to predict the perceptual accuracy of a CC stop cluster from the associated information measures of relevant speech elements in the listeners' languages.

The estimated models show that high informativity has a generally negative effect on the perceptual accuracy of stops. Functional load shows less consistent correlations with perceptual accuracy across different model specifications, but generally has a positive effect on the perceptual accuracy.

In addition, Japanese listeners show highly consistent results across the two experiments and different model specifications. This contrasts with less consistent results for English and Korean listeners.

The current dissertation provides the first empirical evidence for a significant relationship between informativity and functional load, and between the information measures of speech elements and their perceptual accuracy. Furthermore, it reveals how listeners' native

languages affect that relationship, and how the cross-linguistic variation of that relationship may be related to the characteristics of individual languages such as their syllable structures.

To my family

Contents

List of Figures

List of Tables

Acknowledgments

I would like to dedicate this page to express my gratitude to those who encouraged me throughout my PhD study.

I would like to extend my foremost gratitude to my advisor Prof. Keith Johnson for his endless support during my Ph.D study. I would not have been able to finish the PhD program and this thesis without his extreme patience, thorough guidance, and encouragement. I learned much from long hours of discussion with him throughout my research, during which his wide interest and enthusiasm toward Linguistics never ceased to motivate me. I would choose him as my advisor again, if I had to start over the PhD study. I sincerely thank him for making my time in Berkeley as productive as it could be.

I would also like to thank the rest of my dissertation committee: Prof. Sharon Inkelas, Prof. Susan Lin, and Prof. Robert Knight, for their meticulous feedback and insightful comments. Their willingness to have meetings with me and to give suggestion even just to a passing idea guided me to become more creative and widen my perspectives in research. Also, their questions and comments on the thesis made me realize that what I thought I had known was in fact something I did not know well after all. Their guidance helped me bridge the missing links and logical leaps. I benefited much from their efforts and time in making this thesis a better one.

I would also like to thank Prof. Kiyoko Yoneyama at Daito Bunka University, and Prof. Yao Yao at Hong Kong Polytechnic University, who kindly helped me collect data for experiments. I would also like to thank Prof. Hyunkee Ahn at Seoul National University, who was my former advisor, for helping me collect Korean data whenever I needed to. I am grateful for his continuous support. I also appreciate the generosity of Prof. Jiwon Yun at Stony Brook University, who provided me with a valuable resource on Korean phonetics. I thank our IT specialist, Ronald Sprouse, for his immense assistance in resolving a number of technical difficulties. Without their help, the thesis would not have been possible at all.

I am also indebted to a number of people, who helped me survive in Berkeley. I am sincerely grateful to Belen Flores and Paula Floro, who are the most amazing people I have even known. Whenever I needed something, they knew how to get it. Most importantly, they were always kindly helping. I also appreciate the opportunity that Dr. Edward Chang at UCSF offered me, which gave me an access to the Eletrocorticographic data to be used for my qualifying paper. I am also grateful to Samsung Scholarship, which has provided me with a financial support throughout the doctoral program. I would not have been able to focus solely on research had it not been for their generous support.

I thank my fellow (semi-)labmates, Emily Cibelli, Clara Cohen, Sarah Bakst, Gregory Finley, Andrea Davis, Stephanie Shih, Jevon Heath, Matt Faytek, Melinda Fricke, John Sylak-Glassman in the Phonology Lab for all the interesting and stimulating discussions and fun moments in the past years. I also thank other friends in Berkeley, I-hsuan Chen, Elise Stickles, Clare Sandy for all those good times that I shared with them. Without those memories, the graduate school would have been less enjoyable.

Lastly, I would like to dedicate this thesis to my family: My parents, Beom-mo Kang and Hyun-sook Lee, and my brother Hyunmin Kang for their endless care and unconditional love throughout my academic pursuit and my life as a person. Without their prayers, I know that I would not have been able to make it through. I also thank Kyungmin Kim, who endured with me in every step I spent writing this thesis, without whom I am sure that I would have derailed.

Terminology and Abbreviations

Sub-syllabic unit	A speech unit intermediate in size between a phoneme and a syllable (larger than a phoneme and smaller than a syllable). For example, in a CVC syllable, rhyme (VC) and body (CV) are sub-syllabic units.
PI	Phonological informativity. Weighted negative contextual predictability of a speech element.
Universe	A concept from information theory, which defines the set of all possible outcomes or occurrences. In linguistics, a universe is typically the set of all possible speech elements of a certain type. For example, the universe of words is the set of all the possible words within a language.
Entropy	Average negative log probability of the elements in a universe. Used in information theory to represent the amount of uncertainty.
FL (of a contrast between two speech units)	Functional load. The amount of contribution that the contrast makes in differentiating speech elements within a universe (defined above). It is computed as the proportional change in entropy following the removal of the given contrast.
FL (of a speech unit)	Functional load. The amount of contribution that a speech unit makes in differentiating speech elements within the universe (defined above). It is computed as the mean of FL of contrasts between the given speech unit and all of the other speech units.

Chapter 1

Introduction

1.1 Purpose of the study

The relative contribution by different speech elements in transmitting information plausibly varies across different languages, given their phonological differences. As a result, listeners may develop different strategies in attending to various speech elements during speech perception based on their language background. This thesis explores this argument by analyzing the relationship between the information measures of individual speech elements and their perceptual accuracy.

Information measures of speech elements developed from information theory have convincingly accounted for various patterns of speech behavior (Cohen Priva 2008, 2012; Wedel et al. 2013). This thesis provides further evidence on the usefulness of information measures, by documenting the relationship between information measures and perceptual accuracy for the first time. In addition, this thesis shows that different information measures, which emphasize the role of different speech elements in information transmission, vary in terms of the strength and the direction of their relationship with perceptual accuracy. Furthermore, the pattern of this variation is different between the three languages studied, English, Japanese and Korean. Phonological differences across the three languages that may cause these variations are discussed.

Chapter 2 describes the construction of information measures as well as the properties of the corpora of English, Japanese and Korean that are used to compute them. Statistical properties of different information measures and their relationships are also discussed.

Chapters 3 and 4 describe two perceptual experiments, one with audio-only stimuli and the other with audio-visual stimuli. For each experiment, the responses from listeners whose native languages are English, Japanese and Korean are compared with the information measures computed for relevant speech elements in those languages; multiple statistical models are constructed for this comparison. Model estimations with information measures computed at the level of phonemes and other sub-syllabic speech chunks are compared. The results of this comparison vary across the three languages, which is interpreted as reflecting their

different characteristics such as syllabic structures. The findings from the experiments and corpus analyses provide evidence for language-specificity in how the perceptual accuracy for different speech elements is correlated with their information measures.

1.2 Information theoretic framework

This section presents basic concepts of information theory. For the purpose of information theory, information is defined simply as a reduction in uncertainty; it is not related to its broad meaning in everyday life. For example, when someone is listening to a spoken sentence, there is a large uncertainty as to what the whole sentence may be when he has listened only to the first word of the sentence. After he has listened to the second word of the sentence, the uncertainty has become smaller because the second word has become certain; the listener has received some information from the second word. Similarly, when a listener is listening to a word, individual phones and syllables that constitute the word contain information because they reduce the uncertainty about what the whole word may be.

A fundamental concept of information theory is that information can be quantified mathematically. This idea was popularized by Shannon (1948) and Shannon and Weaver (1949). To quantify information, it is necessary to quantify uncertainty, as information is defined as a reduction in uncertainty. Entropy quantifies uncertainty mathematically and represents the amount of uncertainty, or randomness, in a system.

1.2.1 Information-theoretic approaches to speech perception and production

A language is used to transmit a message, and the constituent parts of a message (for example, the words within a sentence or the phonemes within a word) reduce uncertainty on what the whole message may be. By quantifying this uncertainty in a linguistic message, the tools of information theory can be applied to the study of a language.

In the linguistics literature, the information-theoretic approach has been used to explain the usage patterns adopted by the listeners or speakers of a language. While attempting to account for the rate of consonant deletion in American English, Cohen Priva (2008) compares the advantages of frequency- and predictability-based accounts. According to Cohen Priva (2008), studies of consonant deletion in English have found that some stops are deleted more frequently than others; coronals tend to get deleted more often than dorsals or labials. A frequency measure that represents how often a phoneme appears in a given language (Zipf 1929 studies the effect of frequency on speakers' production) can partly explain the deletion patterns of plosives, but fails to do so for nasals. On the other hand, the local predictability account predicts that more predictable sounds would more likely be deleted. However, it is contradicted by some crucial exceptions; for example, [p] in 'workshop' [wɚkʃɒp] is never deleted, even though it is the only sound that can appear after the sequence before the [p] (Cohen Priva 2008, p. 93). In other words, neither the frequency nor the predictability

account fully captures the deletion pattern of stop sounds in general. In search of a better explanation, Cohen Priva (2008) adopts a measure associated with informativity (see section 1.2.2.2).

1.2.2 Information measures of speech units

In information theory, a language is viewed as a sequence of speech elements of a certain size, which can be, for example, phonemes, syllables, words or even phrases, depending on the specific purpose of an analysis. The following subsection presents two information measures that are commonly adopted to estimate the amount of information in a speech unit (as a type, as opposed to a token). First, functional load is introduced with a historical background, and its advantage in accounting for some linguistic phenomena is discussed. Then, a different measure, informativity, which is based on contextual predictability, is introduced.

1.2.2.1 Functional load

According to the functional load (FL) hypothesis (Wedel et al. 2013), the amount of 'work' each phoneme does in differentiating syllables and words is quantifiable. Despite its intuitive appeal, there has never been a clear consensus on the definition of functional load. One way of defining FL at the lexical level, which is adopted in this paper, is to count the number of the minimal pairs attributable to a speech element such as a phoneme. The idea of counting minimal pairs as a measure of FL emerged as early as 1930s (Mathesius 1929, Martinet 1955, King 1967 to Wedel et al., 2013; see also Surendran & Niyogi 2003 for review).

Wedel et al. (2013) categorizes various methods of computing FL, based on the level of analysis (word-level or phoneme-level) and the degree to which an entire system is taken into account; each method implies a different view on the underlying mechanism of a phoneme merger. For each level, based on the assumption on the relationship between the speech unit of interest and the system as a whole, FL measures can represent local- or system-level measures. Following this taxonomy, the most intuitive and well-known method of measuring FL, the minimal-pair approach, represents a local analysis at the word-level. According to the minimal-pair approach, the FL of a phoneme α is the number of word pairs that are differentiated only due to the occurrence of another phoneme in place of α. Another approach, the relative frequency approach, defines the FL of a phoneme as its relative frequency in a corpus, or probability, and is used for a local analysis at the phoneme-level.

In contrast, computing functional load as the change in entropy yields a system-level measure. More specifically, the FL of a phoneme is defined as the proportional change in entropy occurring after every contrast in the system involving the phoneme is removed. This definition of FL is directly adopted from Oh et al. (2013), which follows Hockett's generalization (1967) of system entropy and its change.

Following Surendran and Niyogi (2003), a language, denoted L, is defined as a finite set of all speech elements of a certain size. For example, L can be a set of all words or syllables

within a language. Then, the base entropy, or the amount of inherent uncertainty of the language L is defined by:

Definition 1 *Let N_L be the number of elements in the set L, and let σ_i ($1 \leq i \leq N_L$) denote the elements in L. Let $p(\sigma_i)$ be the probability that the element σ_i occurs. Then, the base entropy, $H(L)$, is:*

$$H(L) = -\sum_{i=1}^{N_L} Pr(\sigma_i) log_2 Pr(\sigma_i). \tag{1.1}$$

A language with a larger set L will tend to have a higher base entropy than a language with a smaller L, given the negative log term, $-log_2 Pr(\sigma_i)$. Similarly, if L is the same, the language for which the possible elements occur with relatively evenly distributed probabilities will tend to have a higher base entropy than the language with relatively uneven distribution of elements.

In addition, each element σ_i (here σ refers to general speech elements, not just syllables) of the language L represents a sequence of sub-elements of a certain size smaller than the size of σ_i itself. For example, if L is the set of words, each of its elements will represent a sequence of syllables or phonemes.

If the contrast between two sub-elements x and y were neutralized, the set L (the inventory of speech elements) would become smaller as some of speech elements σ_i that were distinguished only by the contrast between x and y would become identical (or L remains the same if there were no speech elements in L that were distinguished by the given contrast in the original language). Also, the entropy of the new language either becomes smaller or remains unchanged. With L^*_{xy} denoting the new language after neutralizing the contrast between x and y and $H(L^*_{xy})$ denoting its entropy, the FL of the contrast between x and y can be defined as follows (Surendran and Niyogi 2003):

$$FL(x, y) = \frac{H(L) - H(L^*_{xy})}{H(L)} \tag{1.2}$$

The FL of a single sub-element x is defined by summing $FL(x, y)$ over all possible y (Oh et al. 2013):

$$FL'(x) = \frac{1}{2} \sum_y FL(x, y) \tag{1.3}$$

Functional load has been employed in various linguistic studies. For example, Wedel et al. (2013) finds that two phones of which contrast has high functional load are less likely to merge than those with low functional load. Oh et al. (2013) computes the functional load of individual phonemes for five languages (English, Cantonese, Japanese, Korean and Mandarin) and shows that the relative contribution to the amount of uncertainty within a language by phonological subsystems (consonants, vowels and tones) varies across different languages.

1.2.2.2 Informativity

Another line of literature exploits informativity as an information-theoretic measure (Cohen Priva 2008, 2012 and Seyfarth 2014). Informativity is defined as follows (generalized from Cohen Priva 2012, pp.22-24): Let p denote a speech element and c denote a context. Pr represents probability. With this notation, $Pr(p|c)$ is the probability of the occurrence of the speech element p given context c. The negative log probability, $-logPr(p|c)$, represents how much surprise that a speech element creates under a given context.

$Pr(c|p)$ is the probability of context c given the occurrence of the speech element p. Then, the informativity of p is

$$-\sum_{c} Pr(c|p) log Pr(p|c) \tag{1.4}$$

The expression for informativity can be viewed as the average of $-logPr(p|c)$ over c, where c follows the conditional distribution $Pr(c|p)$. In other words, informativity is the average 'surprise' that a speech element generates across all possible contexts under which it can occur, weighted by the probability of each such context.

The preceding definition of informativity can apply to a speech element of any size. For example, Cohen Priva (2008) defines informativity for individual phones while Seyfarth (2014) uses informativity of words. In this thesis, for convenience, the abbrevation PI, which stands for phonological informativity, is used to denote informativity in general.

1.3 Explaining cross-linguistic differences

In order to interpret cross-linguistic variations in the correlation between information measures and perceptual accuracy, this thesis addresses two alternative concepts that have been frequently referred to by previous studies as factors relevant to perception: Syllable structure and a perceptual unit. Differences in the syllable structure across languages have successfully explained cross-linguistic variations in speech perception and production, which are described in more detail in this section. In addition, there is evidence that perceptual units are language-specific and are determined by the phonological structure of a language. This suggests that information measures computed over speech units of different sizes also may correlate differently for each language.

1.3.1 Syllable structure

Evidence 1: Phonological acquisition of the first language The literature on first language (L1) acquisition provides evidence of a cross-linguistic variation in sensitivity to speech elements of different size and nature, during phonological acquisition of L1. Infants gradually learn to recognize contrasts used for distinguishing speech sounds in their native language from a very early age (Best et al. 1995, Kuhl 2004). Moreover, they also develop

strategies to recognize speech chunks, and to grasp various patterns from an input signal, such as stress feet (e.g. Jusczyk et al. 1993, Jusczyk & Aslin 1995, Jusczyk et al. 1999), which is critical[1] in the segmentation (or parsing) of fluent speech (Nazzi et al. 2006)

In addition, studies have found that different rhythmic properties of languages cause differential development in infants' ability to segment a speech signal (Nazzi et al. 2006, Höhle et al. 2009). Jusczyk et al. (1993, 1999) showed that English-learning infants can segment trochaic bi-syllabic words but not iambic bi-syllabic words, which is consistent with the fact that trochaic stress patterns are more common than iambic stress patterns in English. Closely following Jusczyk et al. (1999), Nazzi et al. (2006) showed that French-learning infants can segment individual syllables, but not iambic bi-syllable words (iamb is acoustically closer to common stress patterns in French than trochee), which reflects the fact that individual syllables themselves, not stress patterns, mostly determine the rhythmic property of speech in French.

The differences in "the size of the segmentation unit" (Nazzi et al. 2006, p.286) that infants exploit prove that they are attuned to speech chunks that represent common speech patterns in their L1. Cross-linguistic studies which compare languages with different rhythmic classes provide evidence that infants' perceptual attention is indeed due to their L1, not due to the language used in creating experiment stimuli (Houston et al. 2000, Polka & Sundara 2012).

In summary, the studies mentioned imply that infants, even from a very early age, experience a cognitive process (of which the underlying mechanism still needs more investigation) that makes them become more sensitive to repeated patterns from their surrounding L1. As a result, they may develop different strategies in assigning their perceptual attention that are optimal for perceiving and learning their L1 in the most efficient way.

Evidence 2: Perception and production of a second language The second line of evidence showing a close relationship between a syllable structure and listeners' perception comes from the vast literature on non-native sound perception. It is well known that listeners perceive a non-native speech signal as its closest match in their L1; whether the foreign speech is a non-native contrast of individual sounds (e.g. r-l distinction by Japanese listeners in Best & Strange 1992 among many); of tones (e.g. Mandarin tonal contrast by the speakers of non-tonal languages, as summarized in a comprehensive review by So & Best 2010); or illicit sound sequences (e.g. word-medial consonant cluster /bz/ by Japanese listeners in Dupoux et al. 1999, and see also literature in Sebastian-Galles 2006). As an example, vowel epenthesis is briefly introduced in this section since it is closely related to the idea that different speech chunks are involved during perception.

Epenthesis in non-native speech perception or production occurs when a string of two consonants that are not attested in L1 (L2 consonant clusters) is heard or produced as the closest match in a listener' L1, which is a string that has a short epenthetic vowel between the

[1]Critical mostly to languages where stress exists. For example, it is important in the segmentation of speech in English, but not in the segmentation of speech in Korean or Japanese.

consonants (Broselow & Finer 1991, Eckman & Iverson 1993, Dupoux et al. 1999, Dehaene-Lambertz, G., Dupoux & Gout 2000, Kabak 2003, Davidson 2006). A number of underlying reasons behind epenthesis have been suggested (see Hall 2011). Among them, the differences in syllable structures between two languages are believed to play an important role (Kabak & Idsardi 2007).

A well-known example of perceptual epenthesis can be found from Dupoux et al. (1999). Dupoux et al. (1999) showed that Japanese listeners hear an epenthetic vowel upon the perception of consonant clusters not present in their L1, such as /bz/. The authors first created a series of non-word sequences of VCuCV (e.g. ebuzo) recorded by a Japanese speaker, from which the vowel /u/ was digitally removed with varying length remaining to make a continuum of increasing duration of the vowel: on one extreme end a vowel was almost completely removed (e.g. eb.zo), and on the opposite extreme, a substantial length of vowel remained (e.g. ebuzo). Japanese and French listeners identified if they heard a vowel from the stimulus and the two groups of listeners showed opposite patterns: While the responses for hearing a vowel remained mostly unchanged irrespective of the actual vowel length for the Japanese listeners, the same responses by the French listeners increased in proportion to the duration of the vowel in the continuum. The supplementary experiments using the stimuli recorded by French speaker confirmed the finding in the first experiment. The result shows that Japanese listeners' perceptual pattern is in line with the way of loanword adaptation in Japanese, where an epenthetic vowel is inserted to resolve the illegal coda (Itô & Mester 1995, cited in Hall 2011) and to have it re-syllabified to be the onset of the following syllable. If a similar re-syllabifying process also operates during perceptual epenthesis, one can assert that CV-dominant Japanese syllables can push listeners to hear illicit clusters to be coherent with their linguistic experience. Jacquemot et al. (2003) conducted an fMRI study to confirm that Japanese listeners' difficulty in distinguishing /ebuzo/ from /ebzo/ is a perceptual phenomenon.

A similar finding on perceptual epenthesis by Kabak and Idsardi (2007) shows that such epenthtic vowel is a result of a syllable structure of listeners' L1. Kabak and Idsardi (2007) focuses specifically on effects of Korean listeners' experience in their L1 syllable on perception of consonant clusters. While replicating the experimental paradigms in Doupoux et al. (1999), Kabak and Idsardi (2007) explicitly investigated the underlying motivations of perceptual epenthesis by testing Korean listeners' perception of two different types of consonant clusters: one that would violate a general well-formedness of a syllable (e.g. /ac.ma/,) and another that would make a permissible sequence (but not necessarily a violation of syllable structure, e.g. /at.ma/). They hypothesized that both types of clusters would suffer from similar degree of mis-perception, if a string of illegal sequence causes a perceptual epenthesis. However, if the knowledge of syllable structure is the motivation of the epenthesis, only the cluster that has a bad coda (which makes a bad syllable) would be mis-perceived by Korean listeners. Perception experiments revealed that Korean listeners heard an epenthetic vowel only for a cluster with the bad coda, supporting the latter hypothesis.

Taken together, both Dupoux et al. (1999) and Kabak & Idsardi (2007) show that phonological knowledge of syllable structure in the case of Japanese and Korean, influences

the process of L2 perception. To reiterate, mishearing 'ebzo' sequence as 'ebuzo' by Japanese listeners indicate their tendency to hear consonant clusters to align with their knowledge of a well-formed CV syllable, the most common syllable types in Japanese. Also, repairing only the bad coda (not bad sequence) with an epenthetic vowel shown by Korean listeners provides another evidence for this tendency.

The effects of L1 syllable structure (as opposed to the knowledge on the string of segments) is further strengthened by findings from the production of non-native consonant clusters by English speakers (Davidson 2006). The production experiment in Davidson (2006) used several types of word-initial consonant clusters recorded by a Czech speaker as stimuli, all of which were not attested in English word-initially. However, the clusters differed in the frequency of occurrences in the English lexicon, since the sequence could occur in other positions in the words (e.g. /ft/ in 'after'). Davidson (2006) correlated production accuracy of each sequence with its frequency measure and hypothesized that they would be positively correlated if the knowledge on the strings of segments was sufficient in processing non-native sound sequence. However, no such correlation was found. Davidson (2006) interpreted the result as speakers' having a higher level (structural) knowledge involving the syllable structure, which may have led them to recruit only the knowledge regarding what can occur word-initially.

In summary, the studies on an epenthetic vowel shows that the syllable structure of listeners' L1 affects both speech production and perception.

Evidence 3: Processing of information in speech Finally, differences in syllable structure and complexity across languages also affect how speakers process information contained in different speech elements.

For example, Pellegrino et al. (2011) found that speakers' speech rate is partly determined by the syllable structure of their L1. The authors compared seven different languages, which varied by the size of syllable inventory and the structure of syllables. In order to measure speech rate, they calculated the average number of syllables per second that are used to convey the same semantic content. Syllabic information density was defined as an average value of semantic information divided by the number of syllables. They found a negative correlation between information density and speech rate. This means that information rate, which is defined as the amount of transmitted message per speech duration, is relatively constant across languages. Their result indicates that the amount of information contained in a syllable affects the speed of its production, which may also affect the speed of its processing.

Differences in syllable structure are also known to affect speakers' recalling rate of a certain speech element. Lee and Goldrick (2008, 2011) particularly looked at the syllable structures of English and Korean by comparing correlation coefficients (a type of associations measure that quantifies strength of the association between two sounds within a speech unit, e.g. a syllable, of a given language, based on the transitional probability) between C and V or between V and C, in each of the two types of sub-syllabic chunks in monosyllables (CV

vs. VC). VC units showed a stronger association than CV units in English, but CV units showed a stronger association in Korean. The authors also investigated if the speakers and listeners of each language had any cognitive bias toward either CVs or VCs in the way that was consistent with the pattern observed in their corpus study. If so, a Korean user would be more sensitive to CVs, but an English user would be more sensitive to VCs. Both Korean and English users were tested with a simple memory task, in which the participants first listened to six pre-recorded monosyllables and recalled them. The result indicated that Korean and English speakers were indeed more sensitive to the sequences observed in the corpus of the their native language. Taken together, Lee and Goldrick (2008, 2011) concluded that the speakers of a language do not just have knowledge of the phonological system of their L1, but exploit the knowledge during speech-related tasks.

Yet, the conclusion is still questionable to some extent, because of the way the result was drawn out from regressions. As previously demonstrated, the correlation coefficient of certain CV and VC sequences was calculated in the set of monosyllables in the respective language and taken as the primary method to quantify syllable structures. For instance, the CV structure with a higher coefficient was considered as the unit associated more strongly than those with a lower coefficient. However, the data from which the coefficient was drawn needs further support from the analysis that actually incorporates the data of usage, since the current coefficient only reflects the existence of CVs and VCs in each corpus. In other words, Lee and Goldrick (2008) took the type frequency of individual sequences, but not their token frequency. The lack of this evidence is particularly problematic, particularly when usage is known to affect speech behavior. The result from Pellegrino et al. (2011) also found that the type of measure that best predicts speakers' pattern is the one that incorporates both the size of inventory (type frequency) and usage of those individual syllables (token frequency).

1.3.2 Perceptual unit

A number of studies in the past attempted to find out whether early stages of speech perception (in particular, before lexical access) operate in terms of a single universal type of speech units. Several studies conducted experiments to measure reaction time to monitoring tasks and argued that syllables, rather than phonemes constituting them, act as the fundamental unit of speech perception. (e.g. Savin & Bever 1970, Foss & Swinney 1973 and Mehler et al. 1981) However, the validity of these experiments have been disputed. (McNeill & Lindig 1973 and Norris & Cutler 1988)

Other studies support the hypothesis that speech units salient in perception are specific to each language. Cutler et al. (1986) finds that French listeners more quickly detect a target syllable when the syllabic segmentation of a stimulus matches that of the target, while English listeners show no such difference. They interpret this result as reflecting the fact that French has clearer and more regular syllable structure than English. Kim et al. (2008) finds that Korean listeners more accurately detect a target syllable when its syllabification matches that of a stimulus, which is attributed to the clear syllable structure of Korean.

Such evidence on language-specific perceptual units suggests the possibility that information measures that more strongly correlate with perceptual accuracy are also partly determined by the size of speech units over which the measures are computed. Therefore, this thesis computes information measures using speech elements of varying sizes and compares their correlation with perceptual accuracy across the three languages, English, Japanese and Korean.

1.4 Current study

This section discusses the contributions of the current thesis and compares them with the findings from the existing literature.

Relationship between information-theoretic measures

The definitions of the two information-theoretic measures, FL and PI, are based on completely different mathematical expressions. PI is defined as the conditional expected value of negative log contextual probability, while FL is defined as a proportional change in entropy following the neutralization of a contrast or a set of contrasts.

However, there is a substantial overlap between their conceptual underpinnings; both measures are system-level measures, computed from the probability distribution of speech elements to capture the amount of information that a speech element carries.

In addition, the two measures successfully correlate with related linguistic phenomena. For example, as discussed before, Cohen Priva (2008) finds that phones with low PI tend to have high deletion rates for American English speakers; and Wedel et al. (2013) finds that phonemes with high FL are less likely to be the target of a sound merger. Although a sound merger and a phone reduction are a diachronic and a synchronic type of change, respectively, there is a link between these two phenomena, particularly because phonetic changes in individual phones contribute to long-term sound changes (Paul 1920, Ohala 1981 and many more, as cited in Garrett & Johnson 2011).

Despite of these conceptual and empirical commonalities between these two measures, the relationship between them has hardly been investigated so far, especially at the sound level. To fill this gap, in Chapter 2, the current work studies how the two measures are correlated.

Effects of PI and FL on perception

Information measures are known to explain some speech behaviors. However, how information measures correlate with perception has not been studied much. This thesis fills this gap by conducting empirical analyses of the statistical relationship between information measures and perceptual accuracy.

Language-specific perception of speech elements

Chapters 3 and 4 find a cross-linguistic variation in how different information measures are corrlated with perceptual accuracy. This cross-linguistic variation is discussed in terms of the different syllable structures of the three languages, English, Japanese and Korean. Syllable structures are known to be related to other speech behaviors. For example, Pellegrino et al. (2011) documents a negative relationship between the rate at which syllables are spoken and the density of information in syllables across different languages. Lee & Goldrick (2008, 2011), as another example, show that English and Korean speakers show differential recalling rates that favor sub-syllabic speech units (CV and VC, respectively) that are more common in their L1.

Chapter 2

Information in Sub-lexical Speech Units: A Cross-linguistic Study

2.1 Chapter introduction

This chapter studies the statistical properties of information measures of different sub-lexical speech elements in English, Korean and Japanese. More specifically, this chapter computes information measures of phonemes and CV and VC units and compares them across the three languages. The results provide background for the analysis of the perception experiments in Chapters 3 and 4.

There is evidence that syllable onset and coda, and sub-syllabic units involving them, have very different statistical properties. Lee and Goldrick (2008, 2011) show that association measures based on transitional probability between C and V within two sub-syllabic entities, CV units and VC units, differ across languages: For English, C and V in VC units are more tightly associated than those in CV units; and for Japanese and Korean, the pattern is the opposite. This variation is attributed to differences in the syllable structure of the three languages. Not only does the number of attested phonemes differ, but the set of sounds allowed in onset or coda is also largely different.

There is also evidence on cross-linguistic variation in the statistical properties of speech elements. For example, Oh et al. (2013) shows that FL of English consonantal phonemes is collectively larger than those in other languages (Korean, Japanese, Mandarin, and Cantonese), due to differences in the phonological system.

In this thesis, two types of information measure quantify the amount of information contained by different sub-lexical units in English, Korean, and Japanese: informativity (generalized from Cohen Priva 2012, pp.22-24; Seyfarth 2014) and functional load (FL; Surendran & Niyogi 2003). Those measures have been frequently adopted to explain different speech behaviors (summarized in Chapter 1). Furthermore, although both PI and FL supposedly represent the amount of transmitted information, no clear link has been found between the two. This chapter specifically addresses this under-studied association between

the two measures and see how PI and FL are correlated empirically.

The investigation of the association between these measures is important, because the quantity of information can be defined in a number of different ways. If the measure of information varies substantially depending on how it is defined, one must be very careful in selecting which measure to use in explaining linguistic phenomena. In other words, if the two measures showed a strong positive correlation, the use of either measure would be acceptable in general. However, if there were zero or even negative correlation between the two measures, each of them should be regarded as representing a particular type of information.

In the sections to follow, the basic properties of the two measures are described, including their definitions. In addition, the construction of a sub-lexical corpus is described for each language; due to the different syllable structures of the three languages, different treatments were necessary.

2.2 Overview

2.2.1 Information measures

The methods of calculating the two information measures, PI and FL, are briefly revisited. To compute them, a syllable corpus is constructed first from an existing linguistic corpus. A syllable corpus is chosen instead of a word corpus or any alternative corpus consisting of sub-lexical speech elements (e.g. a bigram or trigram corpus), because it is relevant to predicting low-level perception data, which is the focus of this thesis. For example, despite of some criticisms on methods of experiments, syllables have long been found to be the perceptual unit most readily accessible to listeners during perception of sub-lexical speech elements (Foss & Swinney 1973, Mehler et al. 1981, and also recent neuroimaging and eletrophysiological evidence reviewed in Hickok & Poeppel 2007). Also, it has been shown that coarticulatory effects between vowels and consonants on perception are greater within syllables than across syllables (as reviewed by Massaro & Oden 1980, pp.133-135). Indeed, it has been found that segments within a syllable are tightly associated, which affects speakers' ability to recall CV and VC chunks upon perception (Lee & Goldrick 2008, 2011).

The use of a syllable corpus implies that in computing the PI of a phoneme, the relevant context is not always the preceding phoneme (as in bigram). For example, in computing the PI of an onset, the context will always follow the onset. There is no reason to assume that speech perception follows a strictly temporal order of onset-vowel-coda, given the continuity of speech signal and the very short time gap between sub-syllabic segments. In addition, tautosyllabic vowels have been found to be an important factor in the perception of consonants, including onsets (Liberman et al. 1967, Diehl et al. 1987).

The corpus used in this study consists of all possible syllables attested in a language, but is different from the list of single-syllable words. For each syllable, the corpus contains information on its frequency of occurrences. Although the corpus data of each language

need a specific treatment prior to the computation of information measures, the same basic methods are used across the three languages.

Following Hockett (1967), Surendran and Niyogi (2003) and Oh et al. (2013), FL is a measure of an entropy change. It was defined in the last section, but is repeated here to accompany the toy example below.

Following the notation introduced in Chapter 1, the base entropy $H(L)$ of L is defined as follows:

$$H(L) = -\sum_{i=1}^{N_L} Pr(\sigma_i) log Pr(\sigma_i). \tag{2.1}$$

The FL of a contrast (x/y) is defined as follows:

$$FL(x,y) = \frac{H(L) - H(L^*_{xy})}{H(L)} \tag{2.2}$$

The FL of a single sub-element x can be expressed as follows:

$$FL'(x) = \frac{1}{2}\sum_{y} FL(x,y) \tag{2.3}$$

A toy example in Table 2.1 illustrates the computation process in more detail. The first column describes the original language L and its elements. In the third column, the contrast between n and l in onset is neutralized. The change can be regarded either as n replaced by l or l replaced by n; the resulting change in entropy is the same. The neutralization in onset reduces the entropy of the language to 2.222 from the original entropy of 2.458. The associated FL is the proportional change in entropy, which is $(2.458 - 2.222)/2.458 = 0.096$. Similarly, the fourth column describes the neutralization of the contrast between n and l in coda.

The second column describes the neutralization of n and l in all syllabic positions, which produces the largest reduction in the number of distinct syllables. Therefore, the entropy of the language decreases the most, and the associated FL is the largest.

Informativity, abbreviated as PI (phonological informativity) in this paper, is computed from contextual probability. The definition used here can be applied to a speech element of any size, and generalizes word informativity from Seyfarth (2014). Its mathematical definition is repeated here to accompany the toy example. With p denoting a speech element of interest and c denoting a context, the PI of p is defined as follows:

$$-\sum_{c} Pr(c|p) log(Pr(p|c)) \tag{2.4}$$

It is worth noting that a context can be defined in different ways. For example, for onset /t/ in the syllable /tap/, just the vowel following /t/, /a/, or the rime, 'ap', can be used as its context.

σ	#	P(σ)	n∼l	#	P(σ)	onset n∼l	#	P(σ)	coda n∼l	#	P(σ)
bal	10	10/43	ba*	15	15/43	bal	10	10/43	ba*	15	15/43
pal	5	5/43	pa*	5	5/43	pal	5	5/43	pa*	5	5/43
ban	5	5/43	–	–	–	ban	5	5/43	–	–	–
bun	10	10/43	bu*	10	10/43	bun	10	10/43	bu*	10	10/43
nup	3	3/43	*up	13	13/43	*up	13	13/43	nup	3	3/43
lup	10	10/43	–	–	–	–	–	–	lup	10	10/43
total	43	1		43	1		43	1		43	1
entropy $H(L)$	$H(L*)$=2.458		$H(L^*)$=1.902			$H(L^*)$=2.222			$H(L^*)$=2.138		
$FL(n,l)$	–		0.226			0.096			0.130		

Note: The neutralization process is shown for one consonantal contrast–/n,l/.

Table 2.1: Toy example of FL computation (adopted and modified from Oh et al. 2013).

Referring back to the toy example in Table 2.1, if the context of an onset is defined as the following vowel, the PI of onset /b/ is computed as follows:

Onset /b/ occurs in syllables 'bal', 'ban' and 'bun'. Therefore, the contexts under which onset /b/ occurs are /a/ and /u/. The conditional probability $Pr(p = [b]|c = [\_a])$ is the sum of the frequencies of 'bal' and 'ban', divided by the sum of the frequencies of the syllables that have a vowel /a/: 'bal', 'pal' and 'ban'. Therefore, $Pr(p = [b]|c = [\_a]) = 3/4$. Similarly, the conditional probability $Pr(p = [b]|c = [\_u])$ is 10/23.

The total frequency of onset /b/ is 25. The frequency of onset /b/ accompanied by vowel /a/ is 15, and the frequency of onset /b/ accompanied by vowel /u/ is 10. Therefore, $Pr(c = [\_a]|p = [b]) = 3/5$ and $Pr(c = [\_u]|p = [b]) = 2/5$.

Therefore, the PI of onset /b/ given following vowel environments is $-Pr(c = [\_a]|p = [b])logPr(p = [b]|c = [\_a]) - Pr(c = [\_u]|p = [b])logPr(p = [b]|c = [\_u]) = 0.5058$.

2.2.2 Method of analysis

PI is computed for onset and coda with the tautosyllabic vowel as the context. The mean of onset PI and the mean of coda PI are compared by two-sample t-tests (except for Japanese; see Section 2.5.3.1 for details). If onset PI is significantly greater or smaller than coda PI, there is a positional bias in information transmission between onset and coda.

With FL, two separate comparisons are made. First, as with the comparison just described for PI, the FL of onset is compared with that of coda. The second comparison is done at a sub-syllabic level: The FL for CV units vs. the FL of VC units, which is motivated by the findings of Lee and Goldrick (2008) on the association of the vowel with the onset and the coda within a syllable. Since the number of possible CV units and VC units is much

Language	Corpus source		Number of words	Part of speech	Morphological complexity
	Frequency	Transcription			
English	Subtlex	CMU, Buckeye	8,400	all (content and function words)	complex words
Japanese	NTT Psych. DB	NTT Psych. DB	65,000	nouns	complex words
Korean	Sejong	Lee (2002)	12,300	content words	complex words

Table 2.2: Summary of corpora used for each language

larger than the number of possible onsets and codas, the sub-syllabic analysis can produce stronger statistical results.

In computing the information measures, onsetless and codaless syllables are coded with a NULL onset and a NULL coda, respectively, and included in the syllable corpus of each language. 'CV','VC', and 'V' syllables are not only phonotactically possible, but also constitute a substantial part of a syllable corpus. Therefore, they are taken into account in computing information measures.

However, NULL onset and NULL coda are not included in the statistical analysis (including correlation test) on the information measures of individual speech elements. This choice has been made to formulate the discussion in terms of actual sounds. In addition, given that the NULL sound is only a single sound among many possible consonants, excluding NULL onset and NULL coda would have little effect on the statistical analysis.

2.2.3 Corpora

Table 2.2 summarizes the original corpus used to construct a syllable corpus for each language. The details of each corpus (e.g. source, selection criteria, etc.) are given in the table.

The source corpus for each language was constructed differently. The Japanese corpus was the most limited in the sense that it contained only nouns. However, reducing the corpora of English and Korean to retain only nouns was not feasible; doing so would reduce the number of words in the English corpus too much, and the resulting corpus would not be representative of the English language (Krishnamurthy 2000, Granath 2007). Therefore, the source corpora were used without any significant modification.

	labial	dental	alveolar	post-alveolar	lateral	palatal	velar	glottal
approximant			ɹ		l	j	w	
fricative	f v	t d	s z	ʃ ʒ				h
affricate				ʧ ʤ				
stop	p b		t d				k g	
nasal	m		n				ŋ	

Table 2.3: English consonants

	Front	Central	Back
High	i, ɪ		u, ʊ
Mid	ɛ	ʌ, (ə), (ɝ)	ɔ
Low	æ		ɑ
Diphthongs		eɪ, oʊ, aɪ, aʊ, ɔɪ	

Table 2.4: English vowels

2.3 Study 1: English

This section presents findings from a corpus analysis of speech units found in the σ-universe (the syllable universe) of English. The probability of occurrences of certain bi-phone sequences has been previously analyzed with monosyllable monomorphemic words. Kessler and Tremain (1997) and a more recent work by Lee (2006) are examples of such studies. Both of these studies investigated the degree of association between segments within each existing monosyllable word, using the frequency of occurrence of a phone in a certain syllabic position and comparing it to the expected frequency predicted by the probability of the phone. The results from the two studies confirmed that the correlation between vowels and codas are much greater than between onsets and vowels. Here, I investigate how such findings are represented by information measures.

2.3.1 Brief summary of phonology

The basic background of English phonology relevant for the analysis is outlined in this section. The consonant and vowel inventories of the English language adopted for this dissertation are illustrated in Table 2.3 and 2.4.

In considering the consonant and vowel inventories of English, the CMU pronunciation dictionary (Weide 1994) as well as the standard textbook description by Ladefoged & Johnson (2015, pp.60-113) were taken into account. The CMU dictionary is a comprehensive documentation of the North American pronunciation of more than 100,000 words, and it has been used as the transcription method to construct a spoken corpus of American English

(Buckeye; Pitt et al. 2007). Since the current work also relies on the CMU dictionary for the purpose of transcription and syllabification (more details on the use of multiple corpora during the construction of a syllable corpus is described in the next section), the consonant and vowel inventory is built using the same coding convention as the CMU dictionary. The CMU pronunciation dictionary marks English phonemes following the transcription code by *Arpabet*, which was developed to transcribe the sounds (or phones) that occur in general American English for speech recognition technology. It is important to note that not all sounds marked by *Arpabet* convention are technically English phonemes, and therefore not implemented in the encoding convention in the CMU dictionary. For example, a flap /ɾ/ and a glottal stop /ʔ/ are coded in *Arpabet*, but are not part of the consonants that the CMU encodes. In other words, allophones of /t/ or /d/ are not reflected in the dictionary.

As for the English vowels, a total of 16 were included in the vowel inventory. Eleven of them were considered monophthongs, and five were considered diphthongs. However, [ʌ] and [ə] are not distinguished in the CMU dictionary.

Among five different R-colored vowels that *Arpabet* could mark, /ɝ/ was considered a vowel element, as opposed to a combination or a simple vowel and coda /ɹ/. This treatment is directly adopted from the conventions of the CMU pronunciation dictionary, where all other rhotacized vowels are transcribed as a sequence of two sounds (e.g. EH R for /ɛɹ/ in 'air'), while /ɝ/ in 'her' is transcribed with only one symbol 'ER'.

2.3.2 Corpus

The frequency of a syllable was computed as the sum of the frequencies of the words in which it occurs. The probability of the occurrence of each syllable was calculated as the ratio of the frequency of the syllable to the sum of the frequencies of all of the syllables ($p(s) = f(s)/sum(f(s))$). The following describes the sources used in constructing the syllable corpus.

List and frequency measures: subtlex-us (Brysbaert & New 2009). The frequency data from word entries in SUBTlex-us corpus (Brysbaert & New 2009) is used as the foundation for lexical frequency measures. The SUBTlex-us corpus is collected from subtitles of US films between 1900 and 2007 and US television series, and consist of 51 million words in total.

Phonetic forms: CMU pronunciation dictionary (Weide 1994). Each word from SUBTlex-us corpus was transcribed using the CMU Pronunciation Dictionary (Weide 1994) into the citation form. This process is crucial in English, where orthography of a word and its actual pronunciation is not necessarily straightforward.

Syllabification of words: ViC Dictionary (Kiesling, Diley & Raymond 2006). In order to create the syllable corpus, each word from SUBTlex-us needed to be syllabified first.

Therefore, another corpus that has information on syllabified words was necessary. For this purpose, a sub-component of the Variation in Conversation (ViC) Project (Kiesling, Diley, & Raymond 2006) was used: The ViC syllabified dictionary (ViC Dictionary). The ViC Project documents how the Buckeye Corpus of Conversational Speech (Pitt et al. 2007) was created in a great detail, and the ViC Dictionary has information on how 10,444 words in the Buckeye corpus should be syllabified along with their pronunciations, based on the CMU encoding.

The syllabification in the dictionary was developed from a version of algorithm by Fisher (1996), which is based on the syllabification rule by Kahn (1976), except for ambisyllabicity. Specifically, the dictionary allows only attested clusters in syllable onset and coda, following a generalization from Kahn (1976, pp.42-43). However, it disallows ambisyllabicity and prefers onset to coda, such that *litter* and *supper* are syllabified as [li.tɚ˞] and [sʌ.pɚ˞] in the ViC dictionary. In summary, the syllabification in the ViC Dictionary adheres to the following rules: First, a syllable's onset is extended as long as it is legal (and only valid onsets can occur); and secondly, ambisyllabic coda is not allowed.

Several errors in the processed (syllabified) words in violation of the two rules were hand-corrected (e.g. [ha.ŋɪŋ] → [haŋ.ɪŋ], because ŋ is not a legal onset). In addition, it is worth mentioning that because of the encoding by the CMU where ɚ˞ is treated as a single vowel, some syllables following this vowel are often transcribed as onsetless, as in [ɪm.tɚ˞.ɑpt].

Interim summary. In summary, a list of words and their frequency measures in SUBTlex-us were matched with their syllabified counterparts from the ViC dictionary, which yielded approximately 8,400 words in total. A new list of every possible attested syllable tabulated from the 8,400 words, which was now marked with syllable boundaries. The frequency of each syllable attested in the SUBTlex corpus (word frequency per million words) was summed from the entire list.

Selection of words. All words (content and function words) in SUBTlex-us that had a corresponding match in the ViC dictionary were selected for the analysis. The words included both morphologically simple and complex words.

Treatment of consonant clusters. For English, consonant clusters presented the biggest challenge in computing PI and FL; in Japanese and Korean, tautosyllabic consonant clusters do not exist. Unlike the previous literature that included only CVC single-syllable words in English (Kessler & Tremain 1997, Lee & Goldrick 2008), all of the entries in SUBTlex that have consonant clusters as syllable margins were incorporated in order to construct a maximally complete σ-universe.

In computing FL, consonant clusters were treated as a sequence of distinct consonants; for example, neutralizing the contrast between /p/ and /t/ would neutralize the contrast between /pr/ and /tr/ as well. However, in computing PI, each consonant cluster was treated as a distinct type of consonant by itself. For example, the cluster /pr/ is not treated as a

sequence of /p/ and /r/, but rather as an additional consonant. Therefore, the probability of /pr/ would not affect the computation of the PI of /p/ in any way.

This treatment is necessary for PI because it would be impossible to define the context as the adjacent vowel otherwise. Also, the definition of the context as the adjacent vowel would become ambiguous without such a treatment; without such a treatment, a consonant could occur without any vowel adjacent to it.

2.3.3 Results

Here, the results from computation of PI and FL from the English σ-universe are reported.

2.3.3.1 Informativity

Onset. Table 2.5 summarizes the PI of each sound in syllable onset in ascending order. Each phoneme is represented with its ARPABET and the corresponding International Phonetic Alphabet (IPA) symbol in the first two columns. As specified in the methods, only the simple onsets are shown for presentation purposes. However, even with the inclusion of complex onsets, simple onsets had lower PIs than most of the complex onsets; only two among the simple onsets, [ʧ] and [ʤ] have a higher PI than a cluster that exhibits relatively low PI for a cluster, such as [fɹ].

There are a couple of interesting observations to note. Among the possible simple onsets, NULL onset (represented as $\emptyset$ in the table) and glide ([j]) are the ones showing the lowest PIs. Recalling that onset PIs were computed over the likeliness of the occurrence of a phone given the context of the following vowel, the low PI measure for NULL onset implies that onset-less syllables are fairly common in English. Glide 'y'[j] had even lower PIs, which aligns with Kessler and Tremain (1997).

The informativity of [ð] is interesting from the view of the markedness theory (Greenberg 1978, Maddieson 1984, as cited in Goldrick 2002). Briefly stated, the markedness theory of sound patterns is a generalization of cross-linguistic regularities, whether or not a presence of sound A implies another type of sound B. According to this theory, fricatives are generally more marked than obstruent stops in word-initial positions (Gamkrelidze 1978). Although there is known association between marked sounds and their actual frequency[1], it is generally assumed that less marked forms are likely to be associated with a higher frequency (Stemberger 1991, Goldrick 2002). Considering this, a relatively low PI measure of the fricative [ð] is surprising, because frequently occurring sounds would normally be of low informativity, and [ð] shows a pattern of high-frequency phones. However, given that all words were included in constructing the syllable corpus, it is reasonable to assume that the English definite article 'the' or pronouns such as 'this' with very high frequencies of occurrence may have contributed to the observation.

[1]In fact, whether more marked forms are used less frequently, or whether the markedness has effects on actual production at all, has been a topic of an active scholarly debate (as reviewed in Goldrick 2002).

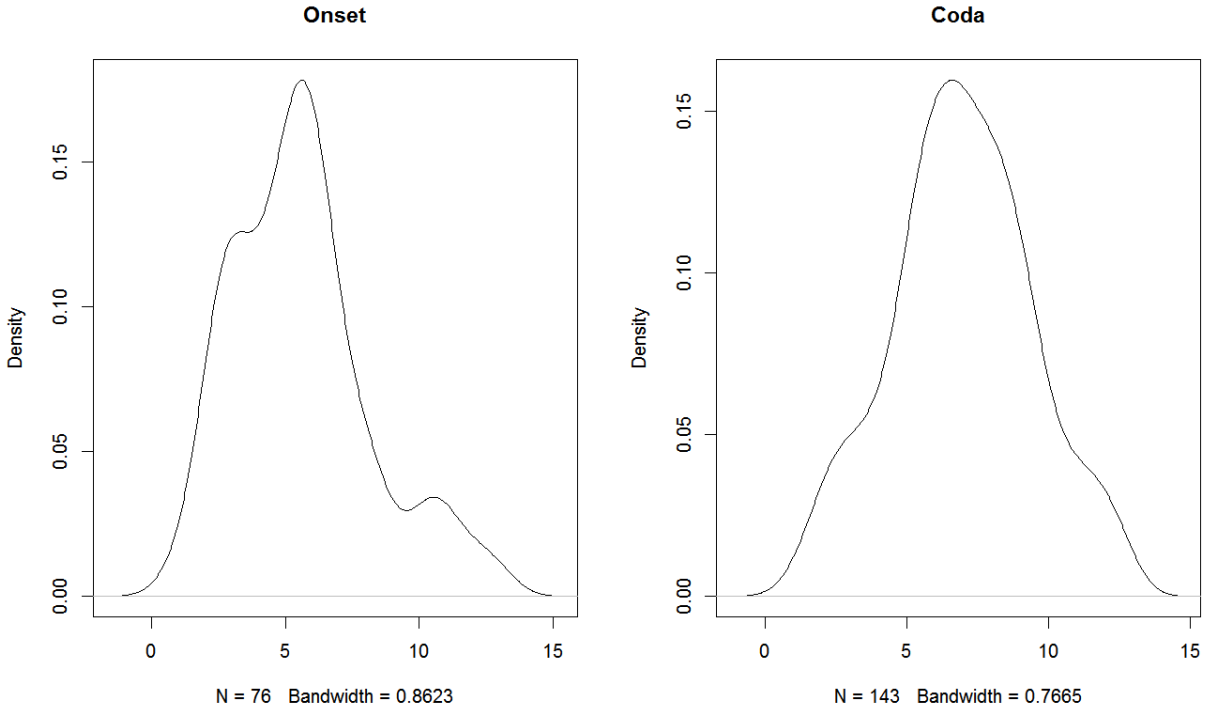

Figure 2.1: Density plot of PI for English syllable onset (left) and coda (right)

Coda. Table 2.6 lists the PI of each phoneme in syllable coda in ascending order. Similar to onset PIs, NULL coda generally showed a lower PI than most of the simple codas, implying that CV syllables are fairly common in English.

Table 2.7 summarizes the descriptive statistics of PI for attested onset and coda, including clusters. There were a greater number of types that occurred as codas than onsets. The density plot in Figure 2.1 for onset and coda illustrates the distributional pattern of PI in more detail.

As shown in Table 2.7, the averaged PI for coda is greater than that for onset. An independent two-sample t-test confirms that the PI for coda is significantly greater than that for onset ($t = 3.8, df = 152, p < 0.01^{**}$).

Simple onset (Arpabet)	IPA	PI
y	j	1.09
∅		1.14
dh	ð	1.78
t	t	2.00
w	w	2.34
h	h	2.52
th	θ	2.57
g	g	2.58
m	m	2.68
n	n	2.72
l	l	2.79
d	d	2.88
b	b	2.88
r	ɹ	2.96
f	f	2.97
s	s	2.99
v	v	3.06
k	k	3.08
jh	ʤ	3.40
sh	ʃ	3.64
p	p	3.75
z	z	4.53
ch	ʧ	4.64
zh	ʒ	5.33

Note: PI is rounded to two decimal places.

Table 2.5: PI of English simple onsets.

Simple coda (Arpabet)	IPA	PI
∅		0.55
ng	ŋ	1.22
r	ɹ	1.85
t	t	1.88
n	n	2.13
s	s	2.22
v	v	2.54
z	z	2.76
d	d	2.77
m	m	2.84
k	k	2.98
l	l	2.99
th	θ	3.59
f	f	3.78
p	p	3.87
ch	ʧ	4.70
g	g	4.87
jh	ʤ	5.22
b	b	5.63
sh	ʃ	5.96
zh	ʒ	7.42
dh	ð	7.67

Note: PI is rounded to two decimal places.

Table 2.6: PI of English simple codas.

	Onset	Coda
N	76	143
min	1.09	1.22
max	12.63	12.43
range	11.54	11.21
sum	424.96	996.8
median	5.43	6.83
mean	5.59	6.97
SE.mean	0.29	0.21
CI.mean.0.95	0.59	0.41
var	6.56	6.12
std.dev	2.56	2.47

Table 2.7: Descriptive statistics of PI for English syllable onset and coda.

2.3.3.2 Functional load

As with PI, NULL elements are presented as ∅. All values were multiplied by 1000 and rounded to two decimal places. FL conceptually represents the amount of system entropy that the σ-universe loses with the loss of a phone x. Therefore, if FL of x was greater than y, x would be more responsible for differentiating words or syllables in the language. Tables in this section list FL in descending order.

Segmental level

Phones across all syllable positions. Table 2.8 shows the computed FL measures for English consonant sounds across all syllabic positions.

The plosive /t/ is the phone with the highest FL measure, meaning that the amount of syllable differentiation that the σ-universe loses is greatest with the loss of contrast between /t/ and other phones. /k/ shows the second highest FL in stop plosives, which is followed by /p/. The order is consistent with the order of phone frequency, previously summarized in Cohen Priva (2008, p.92). Furthermore, the order of FL for nasal stops (/n/>/m/>/ŋ/) and for voiced stops (/d/>/b/>/g/) also follow the phone frequency pattern. The similarity between FL measures and phone frequency is not surprising because FL computation is based on the total frequencies of contrast losses.

Position-specific FL. Table 2.9 and 2.10 show FL of English consonant sounds computed from the removal of phoneme contrasts only in a single syllable position, onset or coda. As both onset and coda, NULL element has the highest FL, meaning that onsetless or codaless syllables are frequent.

The order of the FL measures of the stop plosives in onset and coda is identical to the result found for those calculated across all syllabic position, /t/>/k/>/p/ and /d/>/b/>/g/, conforming to the discussion on the relationship of FL with phone frequency, as mentioned above. However, for nasals, /m/ exhibits a greater FL measure than /n/ does in syllable onset (/ŋ/ is not attested in syllable onset). The pattern suggests that the high FL of /n/ in the analysis on all syllable positions may be attributed mostly to a higher FL of /n/ in coda rather than in onset.

It is noteworthy to mention that /t, d/ were not coded as flap [ɾ], even in the common flapping environment in English (between vowels, where the first vowel is stressed). Although FL of both /t, d/ are comparatively high in the observed data, they might decrease if the two sounds were coded as [ɾ] in the flapping environment; such a change would decrease the frequencies of /t, d/.

The overall descriptive statistics for position specific FL are shown by Table 2.11. Again, NULL onsets and codas are excluded from the analysis. The overall pattern suggests that both the number of onset types as well as their mean is greater for onsets than those for codas. The difference is significant in an independent t-test ($t = 5.2778$, $df = 36.175$, $p<0.001$*).

Sound (Arpabet)	IPA	FL
t	t	54.26
Ø		43.73
m	m	42.35
n	n	41.94
s	s	37.79
d	d	36.72
b	b	29.41
r	r	29.39
l	l	29.12
k	k	27.28
w	w	25.63
dh	ð	25.41
h	h	24.82
f	f	24.45
p	p	23.08
v	v	18.15
g	g	16.59
z	z	16.43
sh	ʃ	15.26
y	j	11.87
ch	ʧ	7.66
jh	ʤ	6.77
th	θ	5.92
ng	ŋ	3.04
zh	ʒ	1.02

Note: All values should be multiplied by 0.001. FL is rounded to two decimal places.

Table 2.8: FL of English consonant across all syllable positions.

Sub-syllabic level. Tables 2.12 and 2.13 show the 50 CV units and VC units with the highest FL, respectively. As expected from the high FLs observed for NULL onsets and codas above, V-only units (onsetless and codaless in CV and VC unit respectively) have high FL.

In addition, the overall descriptive statistics for FLs for the two sub-syllabic units are shown in Table 2.14. The density plot in Figure 2.2 shows the distribution of the FL of CV and VC units. Because the data is highly skewed, log-transformed FL is used for group comparison.

Onset (Arpabet)	IPA	FL
∅		47.31
t	t	41.52
m	m	34.51
d	d	31.89
s	s	30.78
b	b	28.44
w	w	25.63
dh	ð	25.40
n	n	25.23
hh	h	24.82
l	l	23.01
r	r	22.63
k	k	21.51
p	p	20.17
f	f	19.17
g	g	15.42
sh	ʃ	14.51
v	v	13.60
y	j	11.87
z	z	6.83
ch	ʧ	6.29
jh	ʤ	6.02
th	θ	4.59
zh	ʒ	1.00

Note: All values should be multiplied by 0.001. FL is rounded to two decimal places.

Table 2.9: FL of English consonants in syllable onset.

2.3.3.3 Correlation between PI and FL

A Pearson product-moment correlation coefficient between PI and FL measures of each consonant was measured to determine the relationship between the two measures. There is a significant negative correlation between PI and FL ($r = -0.5, t = -3.8, df = 42, p < 0.01$*). Figure 2.3 summarizes how each segment is correlated for PI and FL by different syllabic position. The correlation is significant for both onset and coda (Onset: $r = -0.54, t = -2.9, df = 21, p < 0.01$*; Coda: $r = -0.7, t = -4.2, df = 19, p < 0.01$*). The result

Coda (Arpabet)	IPA	FL
∅		40.14
n	n	16.47
t	t	12.37
z	z	9.56
m	m	7.47
s	s	6.79
r	r	6.64
l	l	5.94
k	k	5.42
f	f	5.16
d	d	4.57
v	v	4.51
ng	ŋ	3.04
p	p	2.83
ch	ʧ	1.37
th	θ	1.33
g	g	1.15
b	b	0.92
sh	ʃ	0.75
jh	ʧ	0.73
zh	ʒ	0.02
dh	ð	0.01

Note: All values should be multiplied by 0.001. FL is rounded to two decimal places.

Table 2.10: FL of English consonants in syllable coda.

differs from correlations of coda in Korean and Japanese, which are explained in more detail in the respective section for Korean and Japanese. In addition, more discussions on the cross-linguistic difference in the correlation of onset and coda are given in Section 2.6.1.

	Onset	Coda
N	27	19
min	1.9	0.73
max	140	16.47
range	138.1	15.74
sum	1187.2	97.01
median	32.4	4.57
mean	44	5.11
SE.mean	7.6	0.97
CI.mean.0.95	15.7	2.04
var	1576.5	17.95
std.dev	39.7	4.24

Table 2.11: Descriptive statistics of FL for English onset and coda

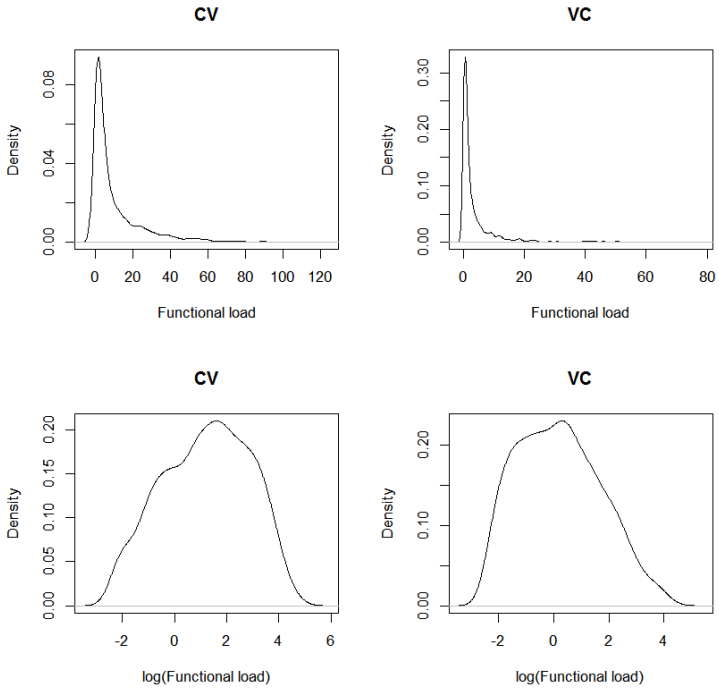

Figure 2.2: Density plot of FL for English sub-syllabic CV unit (left) and VC unit (right).

CV list	IPA	FL	CV list	IPA	FL
yuw1	ˈju	118.24	ow1	ˈoʊ	57.99
dhiy1	ˈði	108.94	ter	tɝ	57.04
ah	ʌ	103.80	shiy1	ˈʃi	56.66
tuw1	ˈtu	103.66	dhey1	ˈðeɪ	55.44
ey1	ˈeɪ	100.80	hher1	ˈhɝ	54.90
miy1	ˈmi	89.41	dher	ðɝ	54.86
tiy1	ˈti	89.04	wey1	ˈweɪ	54.75
eh1	ˈɛ	81.02	er	ɝ	54.12
may	maɪ	78.58	bih	bɪ	53.87
wiy1	ˈwi	77.50	tah	tʌ	53.53
gow1	ˈgoʊ	75.45	wer1	ˈwɝ	53.04
now1	ˈnoʊ	72.20	ah1	ˈʌ	51.50
duw1	ˈdu	72.19	aw1	ˈaʊ	51.35
hhiy1	ˈhi	71.05	geh1	ˈgɛ	51.26
biy1	ˈbi	69.16	sey1	ˈseɪ	50.95
liy	ˈli	65.74	naw1	ˈnaʊ	50.40
aa1	ˈɑ	64.93	ih	ɪ	49.70
ae1	ˈæ	64.58	tuw	tu	49.60
riy	ˈɹi	64.40	diy	di	49.54
ih1	ˈɪ	62.16	diy1	ˈdi	49.09
siy1	ˈsi	60.49	sah	sʌ	48.59
niy	ˈni	59.93	mae1	ˈmæ	48.50
hhae1	ˈhæ	59.42	rih	rɪ	48.30
ver	vɝ	59.16	kah1	ˈkʌ	46.54
sow1	ˈsoʊ	59.13	seh1	ˈsɛ	45.33

Note: All values should be multiplied by 0.001. FL is rounded to two decimal places. The number marks the stress pattern.

Table 2.12: Fifty English CV units with the highest FL.

VC list	IPA	FL	VC list	IPA	FL
iy1	ˈi	112.33	er1	ɜˠ	40.09
ey1	ˈeɪ	103.58	aa1l	ˈɑl	39.64
ah	ʌ	88.88	ahn	ʌn	39.61
ihng	ɪŋ	76.88	eh1m	ˈɛm	37.60
eh1	ɛ	71.27	ae1t	ˈæt	34.89
er	ɜˠ	71.02	ah1n	ˈʌn	33.26
iy	i	67.78	aw1t	ˈaʊt	31.34
ih	ɪ	56.93	eh1n	ˈɛn	30.31
ow1	ˈoʊ	56.64	aw1	ˈaʊ	28.79
aa1	ˈɑ	56.30	eh1r	ˈɛɹ	28.47
ih1t	ˈɪt	56.15	aa1t	ˈɑt	28.08
ih1	ˈɪ	52.87	ih1l	ˈɪl	26.02
ay1	ˈaɪ	52.79	ay1t	ˈaɪt	24.20
ah1	ˈʌ	52.42	ih1r	ˈɪɹ	23.69
eh1s	ˈɛs	50.68	ow1n	ˈoʊn	23.46
aa1n	ˈɑn	50.19	ah1p	ˈʌp	23.14
uw1	ˈu	50.15	ih1f	ˈɪf	22.50
ow1r	ˈoʊɹ	48.41	ow	oʊ	22.44
ae1nd	ˈænd	46.03	ae1z	ˈæz	22.16
ih1n	ˈɪn	45.91	erz	ɜˠz	22.15
ae1n	ˈæn	43.64	ae1k	ˈæk	21.94
aa1r	ˈɑɹ	42.81	ow1ld	ˈoʊld	21.11
ah1v	ˈʌv	41.73	ey1k	ˈeɪk	20.13
ae1	ˈæ	41.38	ahnt	ʌnt	19.50
ih1z	ˈɪz	41.08	eh1d	ˈɛd	18.98

Note: All values should be multiplied by 0.001. FL is rounded to two decimal places. The number marks the stress pattern.

Table 2.13: Fifty English VC units with the highest FL.

	CV	VC
N	967	825
min	0.1	0.091
max	118.24	76.878
range	118.14	76.787
sum	9716.62	3157.217
median	3.67	1.154
mean	10.05	3.827
SE.mean	0.49	0.266
CI.mean.0.95	0.96	0.522
var	229.6	58.311
std.dev	15.15	7.636
coef.var	1.51	1.995

Table 2.14: Descriptive statistics of FL for English CV and VC sub-syllabic units.

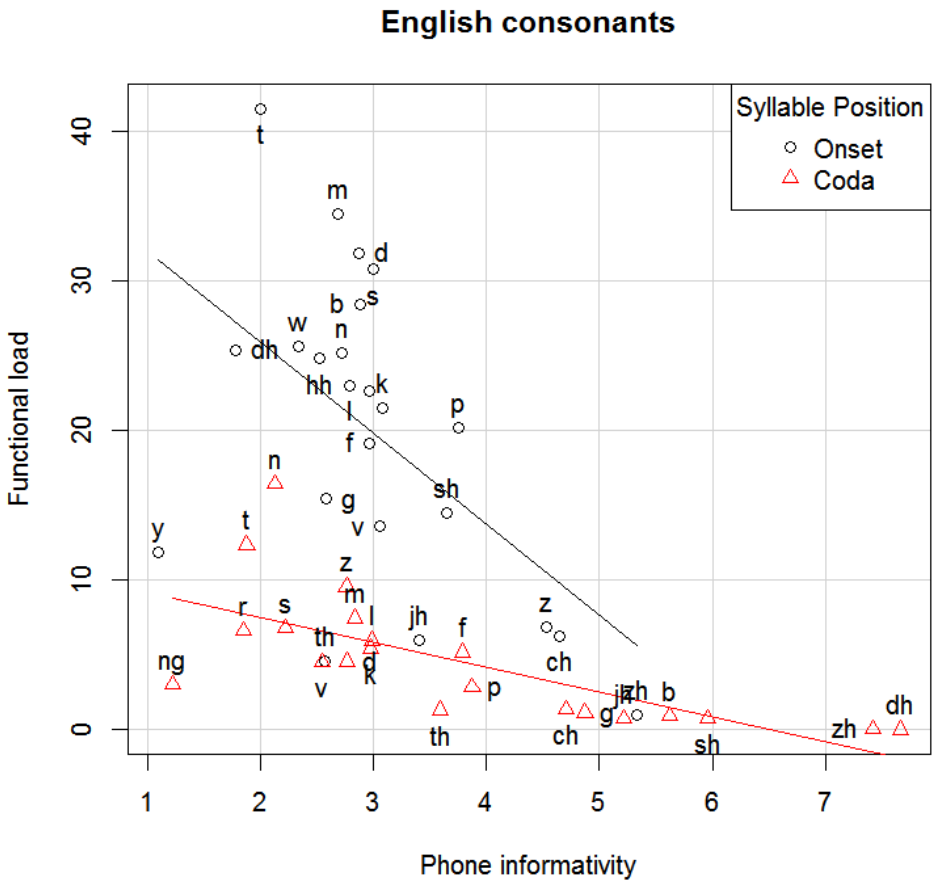

Figure 2.3: Correlation between PI and FL of English consonants by syllabic position.

2.4 Study 2: Korean

This section reports the findings from a corpus analysis of a σ-universe of Korean syllables. As already introduced, Lee (2006) has carried out a rigorous, in-depth investigation on the intrasyllabic associations between segments within monosyllables. The section shows how such findings are represented by information measures.

2.4.1 Brief summary of phonology

The basic background of the Korean language and its phonology relevant to the current analysis are described in this section. In doing so, the overview of the phonetics and phonology of the Korean language by Sohn (1999) and by Kim-Renaud (1974) is followed throughout the section.

The consonant and vowel inventories are illustrated by Tables 2.15 and 2.16 (adopted from Sohn (1999), pp.153-163). In both tables, the phonemes are shown in the standard International Phonetic Alphabet (IPA) symbol. However, standard romanization in Korean phonology is also used when reporting results.

Although still viable among some older speakers, round vowel [y] and [ø] has largely been replaced with diphthong [wi] and [we] (Ahn & Iverson 2007). They were transcribed as 'wi' and 'we' respectively in order to better reflect a more prevalent pronunciation of the vowels in Modern Korean. Besides the 10 simple vowels in Table 2.16, there are eleven additional complex vowels that are produced as diphthongs in the Korean vowel inventory: /ja/, /jə/, /jo/, /ju/, /je/, /jɛ/, /wɛ/, /wa/, /wə/, and /ɨi/.[2]

There are a three-way laryngeal contrast in plosives and a two-way contrast in coronal fricatives in Korean (Sohn 1999).[3] The plain and tense stops are known to be voiced between two vowels (Jun 1994), but this intervocalic stop voicing is not taken into account in the current analysis. The liquid /l/ shows allophonic variation where it is realized as [l] in syllable coda and as [ɾ] in other positions.

Korean syllables can be constructed as the following: V, CV, VC, and CVC (Sohn 1999, p.160) (where V here represents both a simple vowel and a combination of glides and vowels).[4] While all the consonant sounds (plosives, liquid, nasals, fricatives) but /ŋ/ can freely occur in syllable onset, only a limited set of sounds, /p,t,k,m,n, ŋ, l/, can occur in syllable coda. This is due to the way codas are unreleased in Korean (Kim-Renaud 1974, Sohn 1999, Oh 1994), which cause the labial and dorsal stops lose laryngeal contrasts, and are realized to

[2] As mentioned in the inventory in Table 2.15, these diphthong-like vowels are described in this section as the combination of glide and a monophthong. However, in the actual analysis, the diphthongs and glides were also treated as vowels, not a part of a separate consonantal inventory.

[3] Although VOT is known to be the primary cue that distinguishes the three-way contrast, substantial evidence has also found a supplementary role for F0 in the contrast (e.g. Choi 2002, Cho et al. 2002)

[4] As Oh (1994, p.157) states, there is a distinction between underlying forms and phonetic forms in Korean syllables. In underlying syllables, up to two consonants can occur in coda, while in phonetic syllables, only one can be realized. As the scope of the current thesis is speech elements in phonetic syllables, discussion on underlying forms are not further pursued.

		labial dental	alveolar	palatal	velar	glottal
stop	plain	p	t	ʧ	k	
	tense	p’	t’	ʧ’	k’	
	aspirated	p^h	t^h	$ʧ^h$	k^h	
fricative	plain		s			h
	tense		s’			
nasal		m	n		ŋ	
liquid			l			
glide†		w		j		

† Glides are treated as part of complex vowels or diphthongs, not as consonants. They are shown in the consonant inventory for the illustration purpose only.

Table 2.15: Korean consonants (in IPA)

	Front		Back	
	unrounded	rounded	unrounded	rounded
high	i	(y) wi	ɨ	u
mid	e	(ø) we	ə	o
low	ɛ		a	

Table 2.16: Korean simple vowels (in IPA)

their unreleased counterparts [p˺] and [k˺], respectively. And all coronal stops, fricatives and affricates are neutralized to unreleased [t˺]. This phenomenon is termed Coda Neutralization (Sohn 1999, Kim-Reneaud 1974), summarized below:

(2.5) Coda Neutralization (adapted from Sohn 1999, p.165)

$p, p^h, (p') \longrightarrow p˺$

$t, t^h, (t'), s, s', c, ʧ^h, (ʧ'), h \longrightarrow t˺$

$k, k^h, k' \longrightarrow k˺$

;when sounds on the left are produced unreleased before a consonant, a word boundary, or a compound boundary, they lose contrast. Parenthesis indicates that the occurrence of the material inside them are rare and debatable.

2.4.2 Corpus

The frequency of a syllable was computed as the sum of the frequencies of the words in which it occurs. After the syllable list was constructed, the probability of the occurrence of each syllable was calculated as the ratio of the frequency of the syllable and the sum of the frequencies of all of the syllables ($p(s) = f(s)/sum(f(s))$).

Below is a description of the sources used in constructing the syllable corpus.

List and frequency measures: The Sejong Corpus. Words in The Sejong corpus of the National Institute of the Korean Language (The Sejong Corpus) (www.sejong.or.kr) is used as the foundation for the frequency measures. The Sejong Corpus was constructed from first transcribing the recordings collected from various sources, including lectures, informal dialogs, formal dialogs, and news presentations. It consists of approximately one million words that have been tagged for their part of speech.

Phonetic forms: Korean Standard Pronunciation Dictionary (Lee 2002). Each word form from the Sejong Corpus was first matched to its citation form using the Pronunciation Dictionary by Lee (2002). This dictionary (Lee 2002) is the most comprehensive source that documents standard pronunciations in the Korean language and contains phonetic representations of more than 65,000 words.

Syllabification of words. The syllabification was done based on the structure of phonetic syllables of Korean and the rule that CV was always preferred than VC (e.g. *kkoch i* k'och-i [k'o.ʧʰi] 'flower (subject particle)', from Sohn 1999, p.161).

Adhering to this rule, the transcribed words were manually marked with syllable boundaries. Specifically, the syllable boundaries were placed between two vowels (V.V), between two consonants in a cluster (C.C), and between a vowel and a consonant with a following vowel (V.CV) (Sohn 1999, p.160).

Interim summary. In summary, a list of words and their frequency measures in the Sejong corpus were matched with their counterpart phonetic forms from Lee (2002), which yielded approximately 12,300 words in total. A new list of every possible attested syllable was tabulated from the 12,300 words. The frequency of each syllable attested in the Sejong corpus (word frequency per million words) was summed from the entire list.

Selection of words. All words in the Sejong Corpus that had a corresponding match in the Pronunciation Dictionary (Lee 2002) were selected for the analysis. The words included both morphologically simple and complex words. In addition, the included words were all content words, because the dictionary only contained citation forms of those. In other words, the

pronunciation dictionary did not contain citation forms of suffixes, verb endings, particles, or inflected variants of content words forms by themselves.[5]

2.4.3 Results

PI and FL statistics from the Korean σ-universe are reported. The presentation mirrors the format followed in presenting the results for English.

2.4.3.1 Informativity

Onset. Table 2.17 shows PI measures computed for syllable onsets in ascending order. Tense stop /k'/ had the lowest PI measure of all sounds, and the second lowest PI measure is for syllables without onset (NULL onset). That /k/ has a lower PI than NULL onset suggests that onset /k/ is more common than syllables without onsets. The pattern is consistent with the finding in Japanese by Yoneyama (2000), in which /k/ was the most frequent sound.

Also, in the stop category, the triplet of stops (/p, t, k/) in each three-way laryngeal stop series shows a distinct pattern with respect to the PI of the plain, tense, and aspirated stops: For plain stops, /k/ showed the lowest PI then followed by /t/, then /p/(/k/</t/</p/); for tense stops, the order of PI is /t'/ </k'/</p'/; and finally for aspirated stops, the order is, /k^h/ </t^h/</p^h/. For the triplet of the same place of articulation, however, the plain stops always have the lowest PI.

Coda. Table 2.18 shows the PI measures computed for each syllable coda in ascending order. The result demonstrates that PI for NULL codas (in syllable without codas) has the lowest PI among all possible consonant codas, which suggests that CV or V syllables are fairly frequent and common in Korean. Also it is worth mentioning that most of the sonorant codas(/ŋ, n, l/), except for /m/ show lower PIs than obstruent stop codas (/k, t, p/). Such pattern may be explained by regressive nasal assimilation and liquidization that commonly occur in Korean across affixal, compound, and word boundaries (Sohn 1999, p.172)[6] The examples are given below:

(2.6) a Nasal assimilation (Sohn 1999, p.172)
cip-mun 'house gate' [cim.mun]
kkoch-namu 'flower tree' [k'on.na.mu]
hakmun 'learning' [haŋ.mun]

[5]This is not the case for the CMU dictionary, where all instances of inflected forms of a verb, such as *going, goes, go*, are present in the dictionary.

[6]Note that the Korean syllable corpus used for the analysis cannot capture the results of assimilation across word boundaries (in an utterance), since the transcriptions were from individual words from the dictionary. The syllable corpus, however, reflects the assimilation patterns that occur within words including those within morphologically complex words.

Onset	IPA	PI
k	k	1.41
∅		1.52
h	h	1.88
t	t	1.91
c	ʧ	2.08
m	m	2.08
l	ɾ	2.33
s	s	2.49
kh	k^h	2.52
n	n	2.60
p	p	2.75
t'	t'	3.21
ch	$ʧ^h$	3.64
k'	k'	3.95
ph	p^h	4.22
c'	ʧ'	4.31
s'	s	4.34
th	t^h	4.55
p'	p'	5.18

Note: PI is rounded to two decimal places.

Table 2.17: PI of Korean onsets.

b Liquidization (Sohn 1999, p.168)
cinli 'truth' [cil.li]
man-li '10,000 miles' [mal.li]

PIs for coda stops are similar to those for the plain stop onsets (/k/</t/</p/).

Table 2.19 summarizes the overall descriptive statistics for PI for onsets and codas. There are approximately twice as many attested onsets than codas and the averaged PI measure is also slightly greater for onsets than for codas. However, the difference between the PI of onsets and the PI of codas was not reliably different ($t = 0.94, p = 0.35$).

If the distributions of onsets, vowels and codas were uniform and independent, the PI of onsets would be larger than the PI of codas by log[(Number of Onsets)/(Number of Coda)] = 0.410. The observed difference between them is comparable to this number even though it is not significantly different from zero. This result may indicate that the degree of correlation between onset and vowel is similar to the degree of correlation between coda and vowel.

Coda	IPA	PI
∅		0.45
ŋ	ŋ	1.73
n	n	2.11
r	l	2.57
k	k˺	2.83
m	m	2.84
t	t˺	2.88
p	p˺	3.66

Note: PI is rounded to two decimal places.

Table 2.18: PI of Korean codas.

	Onset	Coda
N	18	7
min	1.41	1.73
max	5.18	3.66
range	3.77	1.92
sum	55.45	18.62
median	2.67	2.83
mean	3.08	2.66
SE.mean	0.26	0.23
CI.mean.0.95	0.56	0.57
var	1.25	0.38
std.dev	1.12	0.62
coef.var	0.36	0.23

Table 2.19: Descriptive statistics of PI for Korean syllable onset and coda.

2.4.3.2 Functional load

The results from FL computation are presented in a manner similar to how they were presented for English. As with PI, NULL elements are presented as $\emptyset$. All values are multiplied by 1000 and rounded to two decimal places. The phones are presented in descending order of their FL.

Segmental level

Phones across all syllable positions. Table 2.20 shows the computed FL measure for Korean consonant sounds across all syllabic positions. NULL elements (in onsetless and codaless syllables) are included during the computation and also shown in the table.

NULL elements demonstrate highest FL among the twenty attested phonemes in the σ-universe. The next highest FL is found for coronal nasal /n/. For stops, the relative order of the three phonemes with different places of articulation is the same for all three contrast categories: dorsal > coronal > labial. The pattern is quite different from the well-known markedness hierarchy and thus from the result found in English (coronal > dorsal > labial). However, it is compatible with the frequency count data of Japanese where /k/ was found to be more frequent than /t/ (Yoneyama 2000).

Position-specific FL. Tables 2.21 and 2.22 show the FL of English consonant sounds for two syllable positions, onset and coda, respectively. As demonstrated in both tables, the syllables without explicit onset or coda carry higher FL than other onsets and codas respectively.

With respect to the three-way stop laryngeal contrasts, all plain onsets showed a higher FL than their tense and aspirated counterparts. In addition, three-aspirated stop onsets (/ph/,/th/,/kh/) display almost the lowest FL of all onsets. Within the aspirated stops, the ordering was exactly the opposite of the ordering found in position-independent aspirated stops or codas – i.e. for onsets, the FL for labials is greater than that for coronals, which is greater than that for dorsals (/p^h/>/t^h/>/k^h/), while in position-independent computation, the order is /k^h/>/t^h/>/p^h/. Furthermore, plain stop onsets also share the same ordering that aspirated stops exhibit across all syllable positions.

Such a pattern is interesting, especially when compared to the FL of English stops that are phonetically similar to aspirated stops in Korean. As demonstrated in Section 2.3.3.2, the FL of English onset stops showed a different order (/t/>/k/>/p/).

Sub-syllabic level Tables 2.24 and 2.25 show 50 CV and VC units with the highest FL, respectively. As with English, high FL measures are observed for NULL onsets and codas; and the frequency of occurrences of V-only units are in the list of 50 units with the highest FL, but slightly more in VC (CV:7 out of 50=14%; VC: 14 out of 50=28%)

In addition, the overall descriptive statistics for FLs for the two sub-syllabic units are shown in Table 2.26. Not only are there more types (as indicated by N) in CV- units,

Sound	IPA	FL
∅		173.600
n	n	85.913
k	k	79.634
l	ɾ/l	74.183
c	ʧ	64.149
m	m	61.359
s	s	60.923
h	h	51.296
t	t	51.296
p	p	42.581
ch	ʧʰ	32.414
kh	kʰ	28.076
th	tʰ	27.784
k’	k’	27.319
t’	t’	25.737
ph	pʰ	24.003
c’	ʧ’	19.854
s’	s’	19.205
p’	p’	12.195
ŋ	ŋ	5.560

Note: All values should be multiplied by 0.001. FL is rounded to three decimal places.

Table 2.20: FL of Korean consonant across all syllable positions.

their mean is greater than that of VC units. The standard independent t-test confirms this difference. The finding is consistent with previous findings on sub-syllabic units: CV units have tighter association, which may lead to greater functional load within the syllabic inventories. The density plot in Figure 2.4 more clearly shows the distribution of all CV and VC units as a function of FL. The distribution displayed is slightly skewed and FL measures are log-transformed for group comparison, as it was done for English.

Finally, the leftmost panel in Figure 2.12 illustrates the mean difference of FL for each sub-syllabic unit. A cross-linguistic analysis for any interaction of sub-syllabic unit type and language is carried out and discussed in more detail in Section 2.6.3.

Onset	IPA	FL
∅		100.107
k	k	79.622
c	ʧ	64.147
s	s	60.878
n	n	59.986
r	ɾ	55.126
t	t	51.318
h	h	51.305
m	m	47.513
p	p	42.473
ch	ʧʰ	32.390
k’	k’	27.087
t’	t’	25.689
c’	ʧ’	19.842
s’	s’	19.214
ph	pʰ	14.373
th	tʰ	14.104
p’	p’	12.183
kh	kʰ	11.217

Note: All values should be multiplied by 0.001. FL is rounded to three decimal places.

Table 2.21: FL of Korean consonants in syllable onset.

Coda	IPA	FL
∅		72.426
n	n	25.434
r	l	18.169
k	k˺	16.848
t	t˺	13.691
m	m	12.914
p	p˺	9.587
ŋ	ŋ	5.564

Note: All values should be multiplied by 0.001. FL is rounded to three decimal places.

Table 2.22: FL of Korean consonants in syllable coda.

	Onset	Coda
N	18	7
min	11.22	5.56
max	79.62	25.43
range	68.41	19.87
sum	688.47	102.21
median	37.43	13.69
mean	38.25	14.6
SE.mean	5.01	2.42
CI.mean.0.95	10.57	5.92
var	451.63	40.93
std.dev	21.25	6.4

Note: All values should be multiplied by 0.001. FL is rounded to two decimal places.

Table 2.23: Descriptive statistics of FL for Korean onset and coda.

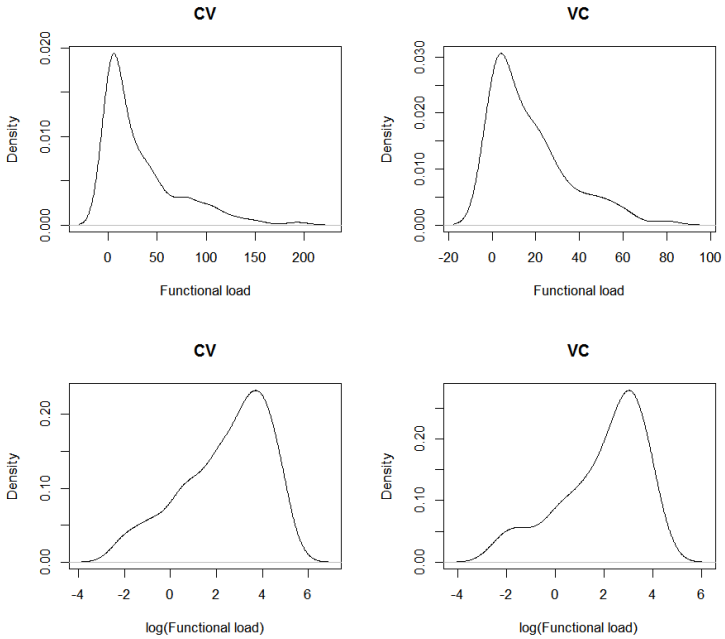

Figure 2.4: Density plot of FL for Korean sub-syllabic CV unit (left) and VC unit (right).

CV list	IPA	FL	CV list	IPA	FL
keu	kɨ	198.21	u	u	98.83
ha	ha	188.34	o	o	96.51
i	i	185.02	mweo	mwə	94.74
keo	kə	158.41	cu	ʧu	93.86
eo	ə	153.59	mu	mu	92.60
a	a	147.28	su	su	88.63
reo	ɾə	146.82	ku	ku	87.32
ka	ka	146.16	pu	pu	86.46
na	na	136.57	k'a	k'a	83.58
ci	ʧi	129.29	rae	ɾɛ	81.64
to	to	127.11	khyeo	k^{h}jə	81.05
ki	ki	123.14	co	ʧo	80.57
ma	ma	122.22	chi	$ʧ^{h}$i	80.20
sa	sa	114.16	kyeo	kjə	79.57
si	si	112.69	tae	tɛ	78.86
ta	ta	110.18	neo	nə	77.71
ca	ʧa	107.41	ko	ko	76.85
ni	ni	106.34	t'ae	t'ɛ	76.02
seo	sə	105.97	nae	nɛ	74.07
ceo	ʧə	103.83	pa	pa	71.63
twe	twe	102.10	hwa	hwa	69.75
ri	ɾi	100.65	reu	ɾɨ	69.44
ce	ʧe	99.34	mo	mo	67.81
po	po	99.15	ye	je	66.41
yeo	jə	99.06	tu	tu	65.62

Note: All values should be multiplied by 0.001
FL is rounded to two decimal places.

Table 2.24: Fifty Korean CV units with the highest FL.

VC list	IPA	FL	VC list	IPA	FL
a	a	164.13	e	e	34.72
i	i	152.87	eop	əp˺	31.14
eo	ə	137.82	eom	əm	30.98
o	o	114.60	ul	ul	30.76
u	u	111.75	ok	ok˺	29.12
ae	ɛ	88.81	eol	əl	28.62
an	an	79.82	eul	ɨl	27.33
eu	ɨ	71.90	we	we	26.14
al	al	62.43	eŋ	eŋ	26.10
eun	ɨn	61.32	ap	ap˺	25.41
eon	ən	58.08	yu	ju	24.92
yeo	jə	54.10	ip	ip˺	24.76
eum	ɨm	53.81	yeol	jəl	24.31
aeŋ	ɛŋ	52.04	eoŋ	əŋ	24.28
ak	ak˺	51.73	on	on	22.85
in	in	48.46	wi	wi	22.46
am	am	47.08	yeŋ	jeŋ	22.32
il	il	44.90	ik	ik˺	21.48
eot	əp˺	43.57	wa	wa	21.15
it	it˺	39.82	yeok	jək˺	20.57
eok	ək˺	39.08	ol	ok	20.39
un	un	38.70	uiŋ	ɨiŋ	19.85
yo	jo	36.76	waŋ	waŋ	19.80
yeon	ən	35.24	im	im	19.47
at	at˺	34.79	oŋ	oŋ	18.86

Note: All values should be multiplied by 0.001
FL is rounded to two decimal places.

Table 2.25: Fifty Korean VC units with the highest FL.

	CV	VC
N	235	97
min	0.11	0.092
max	198.21	79.823
range	198.1	79.731
sum	7621.73	1648.159
median	17.56	11.014
mean	32.43	16.991
SE.mean	2.54	1.785
CI.mean.0.95	5	3.544
var	1513.34	309.181
std.dev	38.9	17.584

Note: All values should be multiplied by 0.001. FL is rounded to two decimal places.

Table 2.26: Descriptive statistics of FL for Korean CV and VC sub-syllabic units.

2.4.3.3 Correlation between PI and FL

A Pearson product-moment correlation coefficient was computed between PI and FL of each consonant in Korean. There is a significant negative correlation between the two measures ($r = -0.54, t = -3, df = 21, p < 0.01^*$). Figure 2.5, which summarizes how each segment is correlated with PI and FL by different syllabic positions in more detail, suggests that the negative correlation between PI and FL of onset segments is much stronger ($r = -0.78, t = -4.7, df = 14, p < 0.01^*$), whereas the correlation for coda is not significant ($r = -0.13, t = -0.29, df = 5, p = 0.77$). More discussion on the cross-linguistic difference in the correlation between onset and coda is given in Section 2.6.1.

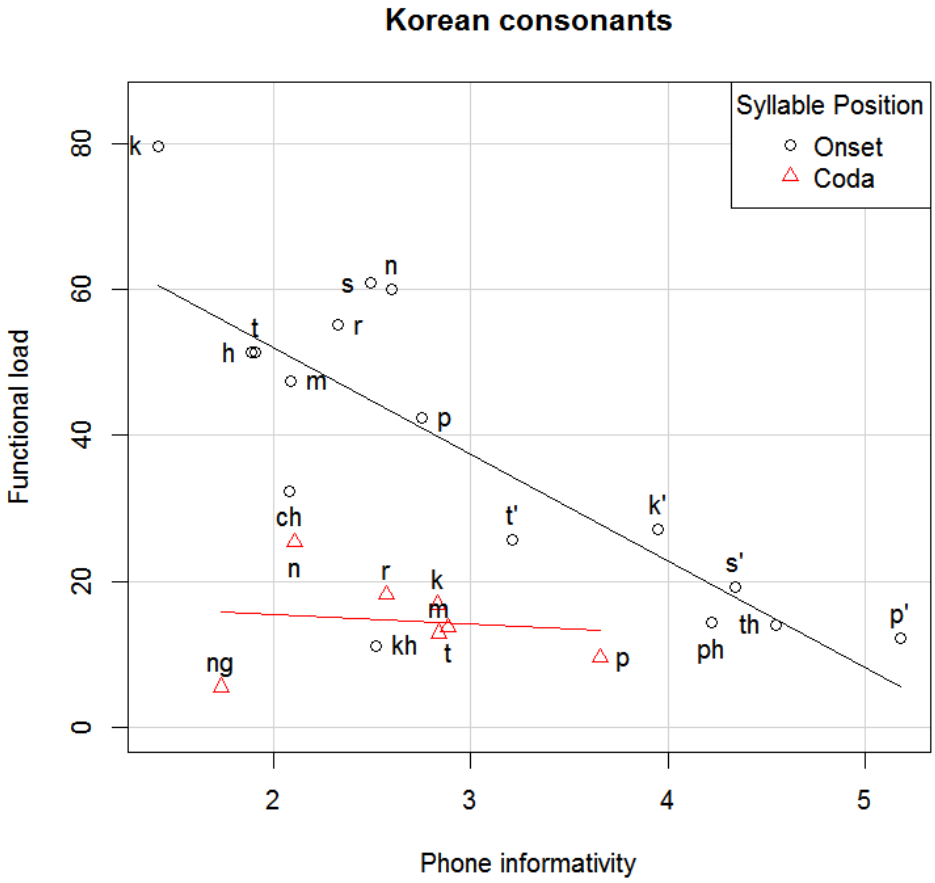

Figure 2.5: Correlation between PI and FL of Korean consonants by syllabic position.

2.5 Study 3: Japanese

This section reports the findings from a corpus analysis of Japanese syllables. The syllable corpus is constructed from the NTT database, which was used in Yoneyama (2000).

2.5.1 Brief summary of phonology

A basic background of the Japanese language and its phonology relevant to the current analysis is described in this section, generally following a recent overview of the phonetics and phonology of the Japanese language by Kubozono (2015a, b).

The phonetic inventory of consonants in Modern Japanese is shown in Table 2.27 (adapted from a combination of Shibatani, 1990:159, Tsujimura 2013:7 , Kubozono 2015:7).[7]. The vowel inventory is shown in 2.28 (adapted from Kubozono (2015a), pp.2-6). In both tables, the phonemes are shown in the standard International Phonetic Alphabet (IPA) symbol. However, standard romanization in common in literature on Japanese phonology is also used when reporting results.

	labial	alveolar	postalveolar	palatal	velar	glottal
plosive	p, b	t,d			k g	
fricative	ɸ	s, z	ʃ			h
affricate		ts	ʧ, ʤ			
nasal	m	n				ŋ
liquid				ɾ		
glide	w			j		

Table 2.27: Japanese consonants (in IPA)

	Front	Central	Back
high	i, iː		ɯ, ɯː
mid	e, eː		o, oː
low		a, aː	

Table 2.28: Japanese vowels (in IPA)

There are 20 consonants and 5 simple vowels. The vowels also have a length contrast, and therefore each of the five simple vowels has a long counterpart, resulting in 10 vowels in total. There may be three diphthong vowels (/ai/, /oi/, /ui/) in Modern Japanese (Kubozono 2015a, p.6). However, the phonetic representations in the NTT Database (details of the

[7]Some of the consonants are debatable as to their phonemic status due to the restrictions on the occurrence and the origin of words in which they occur (Pintér 2015, p.123). However, for the purpose of the current analysis, the phonetic inventory that appears as standard textbook examples (e.g. Tsujimura 2013, Shibatani 1990) is employed

corpus shown below) did not have any instance of such diphthong, possibly due to the debatable status of those vowels; therefore they are not included in the analysis of the current thesis.

With the exception of two moraic consonants, Japanese syllables do not allow coda. The moraic consonants are realized with the place of articulation (POA) of the following consonant (Kubonozo 2015a, pp.9-10). The examples adapted from Kubozono (2015a) below specifically show how syllable coda follows the POA of the following obstruents and nasals.

(2.7) a [ip.pa.i] 'one cup'
[it.ta.i] 'one body'
[ik.ka.i] 'one time'

b [am.ma] 'massage'
[an.na] 'Anna (a girl's name)
[maŋ.ga] 'cartoon'

(adapted from Kubozono 2015a, p.10)

No consonant cluster occurs as onset in Japanese syllables. The entity closest to an onset consonant cluster would be the combination of a consonant and the glide [j], which results in a palatalized consonant. However, the exact structure of palatalized consonants (whether they should be considered as a single onset or a complex onset with two consonants inside) is itself controversial (Kubozono 2015b, p.313). In the current analysis, the palatalized consonants are considered as a single onset as it greatly simplifies the analysis of Japanese syllables (Vance 2008, p.230). Therefore, the following are included as onsets: py, by, my, ny, ry, ky, gy, hy [p^j, b^j, m^j, n^j, r^j, k^j, g^j, h^j].

In summary, the syllable structure of the Japanese language is assumed as (C)V(C), where the parenthesis indicates optionality (Kubozono 2015a, p.16). Specifically, in (C)V(C), V can be either short or long vowel and the onset C can be simple or palatalized consonants, and coda can be [p, t, k, m, n, ŋ], depending on the following consonant.[8]

2.5.2 Corpus

A syllable corpus consisting of all of the possible syllables attested in the lexical corpus along with their frequency was constructed. The frequency of a syllable was computed as the sum of the frequencies of the words in which it occurs. After the syllable list was constructed, the probability of the occurrence of each syllable was calculated as the ratio of the frequency of the syllable and the sum of the frequencies of all of the syllables ($p(s) = f(s)/sum(f(s))$).

[8]The constraint involving syllable weight prevents long vowels and syllable coda to co-occur to make a "superheavy syllable" (Kubozono 2015a, p.16). However, the superheavy syllables occur in some loanwords (e.g. rinkaan 'Lincoln' from Kubozono 2015a, p.13). They are sometimes analyzed as a combination of two syllables (e.g. rin.ka.an/ *rin.kaan, *ibid.*), but syllable weights and related issues such as mora are not further discussed in the analysis as they are beyond the scope of the current thesis. In the NTT Database, they are treated as a sequence of a long vowel and a nasal.

The following describes the sources used in constructing the syllable corpus.

List and frequency measures: The NTT Database (Amano & Kondo 1999, 2000). The NTT Psycholinguistic Database (Kondo & Amano 1999, 2000) was used as the basis on which the syllable corpus was generated. The NTT Psycholinguistic Database (The NTT Database) is the most comprehensive source of the Japanese language with a very detailed information on lexical properties, such as word familiarity, occurrence of frequency, syllable type, phonetic forms etc. Frequency measures were calculated from Asahi newspapers published between 1985 and 1998, resulting in over 300,000 different words types. However, phonetic representations were available only for nouns. Therefore, the set of all nouns, which numbered around 65,000, were selected for the current study.

Phonetic forms: The NTT Database. The nouns in the NTT Database contains phonetic representation of the nouns as well as the syllable boundary of each word, therefore no additional process of matching to the pronunciation dictionary was necessary. However, there were two entities, moraic sounds, that were not in the form of phonetic representations. Those sounds needed manual treatments, which is described in more detail below.

Syllabification of words. Each noun in the NTT database has its corresponding *nearly*-phonetic representation including syllable boundary. The reason that it is nearly phonetic is that the transcription maintains the two moraic consonants, moraic obstruent and moraic nasal. Therefore, an additional treatment was done to these two consonants.

They were treated as syllable coda and recoded to the sounds that are close to the actual realization of sounds. The moraic obstruents and moraic nasals were transcribed to match the POA of the following obstruent and nasal, respectively, which is illustrated in the example given in (2.7). The syllable boundary was placed between two consonants (VC.CV).

There were some ambiguous contexts where it was unclear how the moraic obstruents should be actually realized in a word-final position, which were due to transcription errors. The total number of such occurrences were very infrequent and not included in the analysis. In addition, moraic obstruents before fricatives were recoded to reflect the germination of fricatives (e.g. aQsa → as.sa, where Q represents the moraic obstruent), following Kawahara (2015).

The occurrence of voiced geminate obstruents were extremely rare (occurred in only of < 0.0036 % of the nouns in the original NTT DB), because they are used exclusively for loandwords (Kubozono et al. 2008). Those sounds were transcribed to their voiceless counterparts (e.g. eg.gu → ek.gu).

Interim summary. A new list of every possible attested syllable was tabulated from the 65,000 words marked with syllable boundaries. The frequency of each syllable was computed as the sum of the frequencies of the words in which it occurs.

Selection of words. All 65,000 nouns in the NTT Database that had an associated phonetic representation were selected for the analysis. The words included both morphologically simple and complex words.

2.5.3 Result

Statistical properties of PI and FL computed from the syllable corpus are reported in this section. They are presented in a manner similar to how they were reported for English and Korean.

2.5.3.1 Informativity

Onset. Table 2.29 shows the distribution of the PI of consonants in the syllable onset position in ascending order. Compared to English and Korean, PI for NULL onset (represented as $\emptyset$) is relatively high compared to other onset consonants, which suggests that Japanese syllables generally occur with onsets. To a certain extent, the finding is consistent with Yoneyama (2000), which showed that Japanese syllables generally begin with a consonant.

The ordering of the PI measures for the three stops differing in place of articulation indicates the following order: For voiceless stops, /t/ shows the lowest PI[9], followed by /k/, then by /p/(/t/ </k/</p/); and for voiced stops, PI of /g/ is lower than /d/, which is lower than /b/ (/g/ </d/</b/). In both voicing categories, labials show the highest PI measures of the three-stop triplets, indicating that bilabial stops are more informative than coronal and dorsal stops. In addition, onset [t] shows the lowest PI measures of all onsets.

Coda. Table 2.30 shows the PI measures computed for all eight possible syllable codas (including NULL) in Japanese in an ascending order. As in Korean, the PI for NULL codas (in syllables without codas) was lower than the PI for the other possible codas, suggesting that a syllable without coda is fairly common in Japanese. This observation is compatible with the finding on the prevalence of CV syllables in Yoneyama (2000).

Table 2.31 presents the descriptive statistics of PI for onset and coda.

[9] /t, d/ become /ʧ, ʤ/ before vowel /i/ and /t/ becomes /ts/ before /ɯ/. Such affrication process is reflected in the phonetic transcription.

Onset	IPA	PI
t	t	1.01
h	h	1.13
n	n	1.22
g	g	1.83
ts	ts	1.89
k	k	2.01
∅		2.29
s	s	2.76
sh	ʃ	2.80
d	d	2.84
m	m	2.87
f	ɸ	3.23
r	ɾ	3.25
ch	ʧ	3.34
j	ʤ	3.37
y	j	3.61
ky	k^j	3.70
gy	g^j	3.88
hy	h^j	4.00
b	b	4.06
ry	$ɾ^j$	4.30
z	z	4.37
w	w	4.50
ny	n^j	4.67
py	p^j	5.02
by	b^j	5.13
p	p	5.19
my	m^j	6.28

Note: PI is rounded to two decimal places.

Table 2.29: PI of Japanese onsets.

Coda	IPA	PI
∅		0.06
n	n	2.75
s	s	4.49
ŋ	ŋ	5.55
k	k	5.72
m	m	5.98
p	p	6.20
t	t	9.76
f	ɸ	13.27

Note: PI is rounded to two decimal places.

Table 2.30: PI of Japanese codas.

	Onset	Coda
N	27	8
min	1.01	2.75
max	6.28	9.76
range	5.27	7.01
sum	92.26	46.34
median	3.37	5.80
mean	3.42	5.79
SE.mean	0.26	0.69
CI.mean.0.95	0.53	1.64
var	1.79	3.85
std.dev	1.34	1.96
coef.var	0.39	0.34

Note: PI is rounded to two decimal places.

Table 2.31: Descriptive statistics of PI for Japanese syllable onset

2.5.3.2 Functional load

As with PI, NULL elements are presented as ∅. All FL values were multiplied by 1000 and rounded to two decimal places. The phones are sorted in a descending order of their FL.

Segmental level

Phones across all syllable positions. Table 2.32 shows the FL for Japanese consonant sounds computed across all syllable positions. NULL elements (in onsetless and codaless syllables) were included during the computation and are also shown in the table for the purpose of illustration (not analyzed in the descriptive statistics).

The voiceless velar stop /k/ exhibits the highest FL measure of all the sounds across all syllable positions. Such a high FL for /k/ implies that onset /k/ differentiates the greatest number of syllables in Japanese, which is compatible with the frequency count data of Japanese, where /k/ was found to be more frequent than /t/ in terms of segment frequency (Yoneyama 2000).

The relative order of labial, coronal, and dorsal stops in FL (dorsal > coronal > labial) is similar to that in PI.

Position-specific FL. Tables 2.33 and 2.34 show the FL of Japanese consonant sounds specific in syllable onset and coda, respectively. The FL of NULL onset is lower than that of /k/, /h/, /g/, /n/, and /t/. It is different from the observations from English or Korean, in which NULL onset always showed the highest FL of all other onsets. This pattern again can be accounted by the prevalence of CV syllables in Japanese (Yoneyama 2000). The highest FL measure shown by the NULL coda is also in line with this observation.

Table 2.38 summarizes the descriptive statistics of PI for onset and coda.

Sound	IPA	FL
k	k	142.64
∅		123.47
h	h	111.31
n	n	104.95
g	g	99.44
t	t	98.35
m	m	81.39
r	ɾ	81.15
s	s	65.17
sh	ʃ	63.52
y	j	50.11
d	d	47.23
b	b	43.04
j	ʤ	36.15
ch	ʧ	32.40
z	z	28.79
ky	k^j	20.86
p	p	17.69
w	w	16.31
ry	$ɾ^h$	14.10
ts	ts	13.41
f	ɸ	10.04
gy	g^j	7.94
ny	n^j	5.60
hy	h^j	5.32
ŋ	ŋ	4.55
py	p^j	2.31
by	b^j	2.31
my	m^j	1.89

Note: All values should be multiplied by 0.001. FL is rounded to two decimal places.

Table 2.32: FL of Japanese consonant sounds across all syllable positions.

Sub-syllabic level. Tables 2.36 and 2.37 show the 50 CV and VC units with the highest FL, respectively.

In addition, the overall descriptive statistics for FL for the two sub-syllabic units are shown in Table 2.38. Not only are there more types (as indicated by N) in CV units, its mean

Onset	IPA	FL
k	k	139.98
h	h	111.31
g	g	99.44
n	n	98.53
t	t	96.44
∅		95.46
r	ɾ	81.15
m	m	79.50
sh	ʃ	63.52
s	s	63.38
y	j	50.11
d	d	47.23
b	b	43.04
j	ʤ	36.15
ch	ʧ	32.40
z	z	28.79
ky	kʲ	20.86
w	w	16.31
p	p	16.19
ry	ɾʲ	14.10
ts	ts	13.41
f	ɸ	10.04
gy	gʲ	7.94
ny	nʲ	5.60
hy	hʲ	5.32
py	pʲ	2.31
by	bʲ	2.31
my	mʲ	1.89

Note: All values should be multiplied by 0.001. FL is rounded to two decimal places.

Table 2.33: FL of Japanese sounds in syllable onset.

is greater than that of VC units. The standard independent t-test confirms the significance of this difference. The finding is consistent with what has been previously found for Japanese sub-syllabic units in Lee and Goldrick (2011): CV units have a tighter association.

The density plot in Figure 2.6 more clearly shows the distribution of the FL of all CV and

Coda	IPA	FL
∅		27.76
n	n	6.38
ŋ	ŋ	4.55
k	k	2.65
t	t	1.91
m	m	1.88
s	s	1.78
p	p	1.48
f	ɸ	<0.01

Note: All values should be multiplied by 0.001. FL is rounded to two decimal places.

Table 2.34: FL of Japanese consonants in coda

VC units. The distribution is slightly skewed, and the FL measures were log-transformed for group comparison, as previously done for English and Korean.

Finally, the rightmost panel in Figure 2.12 illustrates the mean difference in FL for each sub-syllabic unit. A cross-linguistic analysis for any interaction of sub-syllabic unit type and language is discussed in more detail in 2.6.3.

	Onset	Coda
N	27	8
min	1.9	0.034
max	140	6.38
range	138.1	6.34
sum	1187.2	20.668
median	32.4	1.90
mean	44	2.58
SE.mean	7.6	0.70
CI.mean.0.95	15.7	1.66
var	1576.5	3.94
std.dev	39.7	1.99

Note: All values should be multiplied by 0.001. FL is rounded to two decimal places.

Table 2.35: Descriptive statistics of FL for Japanese onset and coda.

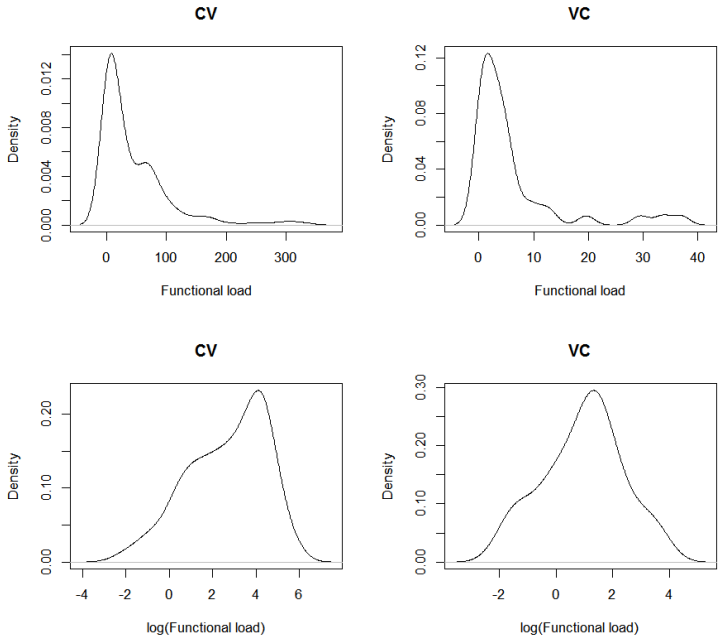

Figure 2.6: Density plot for FL for Japanese sub-syllabic CV unit (left) and VC unit (right).

CV list	IPA	FL	CV list	IPA	FL
ha	ha	333.19	se:	se:	95.64
to	to	309.78	na	na	94.20
ni	ni	299.97	mi	mi	91.99
ga	ga	275.28	hi	hi	91.45
ka	ka	241.70	da	da	90.58
i	i	211.67	a	a	86.91
no	no	205.33	ba	ba	85.20
mo	mo	178.39	o	o	83.28
ku	kɯ	172.12	sho:	ʃo:	79.47
ko	ko	170.46	me	me	79.46
shi	ʃi	163.71	se	se	79.13
ra	ɾa	153.90	shu	ʃɯ	78.88
ki	ki	144.31	wa	wa	78.30
ta	ta	140.38	e	e	78.19
de	de	129.10	ro	ɾo	77.42
tsu	tsɯ	128.90	yo	jo	76.67
u	ɯ	125.68	ke	ke	75.87
ri	ɾi	118.77	go	go	75.16
sa	sa	114.83	su	sɯ	74.52
ji	ʤi	107.91	fu	ɸɯ	74.12
te	te	106.86	yo:	jo:	73.29
chi	ʧi	106.78	to:	to:	71.72
ya	ja	106.07	he	he	70.87
ma	ma	101.11	ne	ne	69.30
ko:	ko:	96.93	do	do	68.93

Note: All values should be multiplied by 0.001. FL is rounded to two decimal places. The colon(:) marks the length contrast.

Table 2.36: Fifty Japanese CV units with the highest FL.

VC list	IPA	FL	VC list	IPA	FL
a	a	113.74	em	em	4.41
o	o	106.44	am	am	4.00
i	i	90.55	it	it	3.79
u	ɯ	67.44	as	as	3.54
o:	o:	60.64	es	es	3.13
e	e	58.28	is	is	2.96
en	en	36.91	ap	ap	2.84
e:	e:	35.84	at	at	2.57
an	an	33.51	im	im	2.53
in	in	29.45	uŋ	ɯŋ	1.96
on	on	19.58	ot	ot	1.82
eŋ	eŋ	13.12	us	ɯs	1.60
aŋ	aŋ	11.79	om	om	1.48
u:	ɯ:	10.69	uk	ɯk	1.08
iŋ	iŋ	9.40	op	op	1.01
un	ɯn	9.07	up	ɯp	0.98
a:	a:	8.05	os	os	0.91
ak	ak	6.42	um	ɯm	0.64
oŋ	oŋ	6.13	ut	ɯt	0.55
i:	i:	6.05	ep	ep	0.54
ip	ip	5.66	e:n	e:n	0.28
ok	ok	4.89	i:n	i:n	0.27
et	et	4.77	o:n	o:n	0.23
ek	ek	4.70	a:n	a:n	0.21
ik	ik	4.49	i:m	i:m	0.06

Note: All values should be multiplied by 0.001. FL is rounded to two decimal places. The colon(:) marks the length contrast.

Table 2.37: Fifty Japanese VC units with the highest FL.

	CV	VC
N	194	40
min	0.09	0.15
max	333.19	36.91
range	333.10	36.76
sum	8456.30	243.37
median	21.30	3.05
mean	43.59	6.08
SE.mean	4.23	1.4
CI.mean.0.95	8.34	2.83
var	3467.10	78.37
std.dev	58.88	8.85

Note: All values should be multiplied by 0.001. FL is rounded to two decimal places.

Table 2.38: Descriptive statistics for FL of Japanese CV and VC sub-syllabic unit

2.5.3.3 Correlation between PI and FL

To study the relationship between the two measures, a Pearson product-moment correlation coefficient was computed between PI and FL. There was a significant negative correlation between the two measures ($r = -0.72, t = -5.8, df = 32, p < 0.01^*$), when coda and onset were both included. Figure 2.7 illustrates how the PI of each segment is correlated with FL by different syllabic position. Figure 2.7 also suggests that the negative correlation between PI and FL (of segments) is stronger for onsets than for codas, which is confirmed by a Pearson's r that reaches significance with onsets ($r = -0.79, t = -6.47, df = 25, p < 0.01^*$), but not for coda ($r = -0.61, t = -1.71, p = n.s.$).

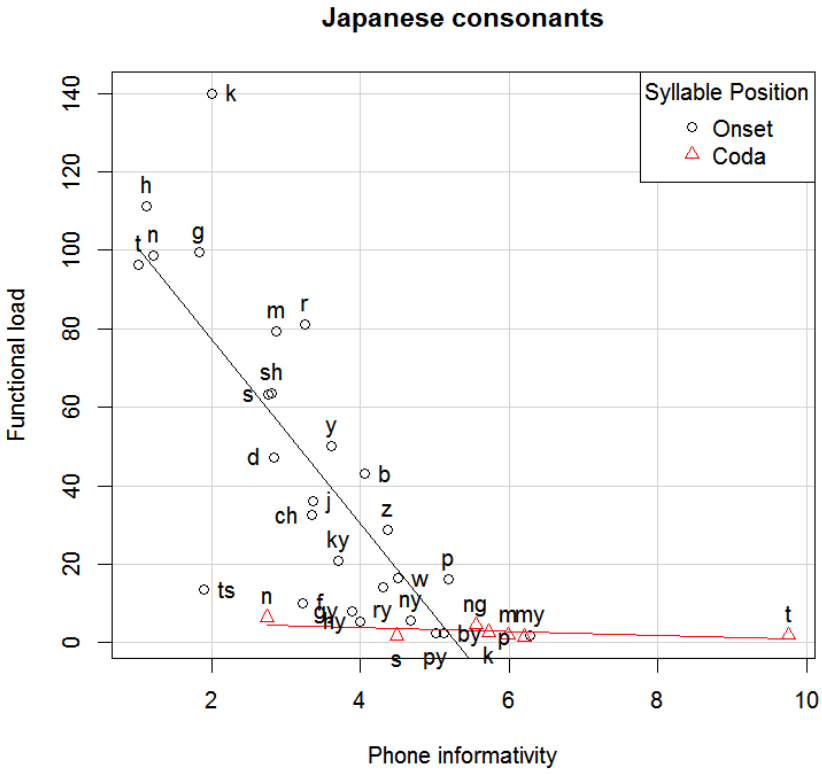

Figure 2.7: PI and FL of Japanese consonants.

2.6 Discussion

2.6.1 Correlations between informativity and functional load

The correlations between PI and FL were consistent across English, Korean, and Japanese. In particular, the PI and FL of onsets are negatively correlated in all the three languages, as shown in Figure 2.8. In contrast, there is no significant correlation between the PI and FL of codas in Japanese and Korean; only English codas showed a negative correlation. However, an additional analysis reveals that the PI and FL of the codas that are common to Korean and Japanese (/m, n, ŋ, p, t, k/), which constitute almost all of the codas in the two languages, show no significant correlation, even in English (shown in 2.9). However, a negative correlation still exists between the PI and FL of all the other codas in English. This result suggests that the correlation between PI and FL may vary across different types of codas.

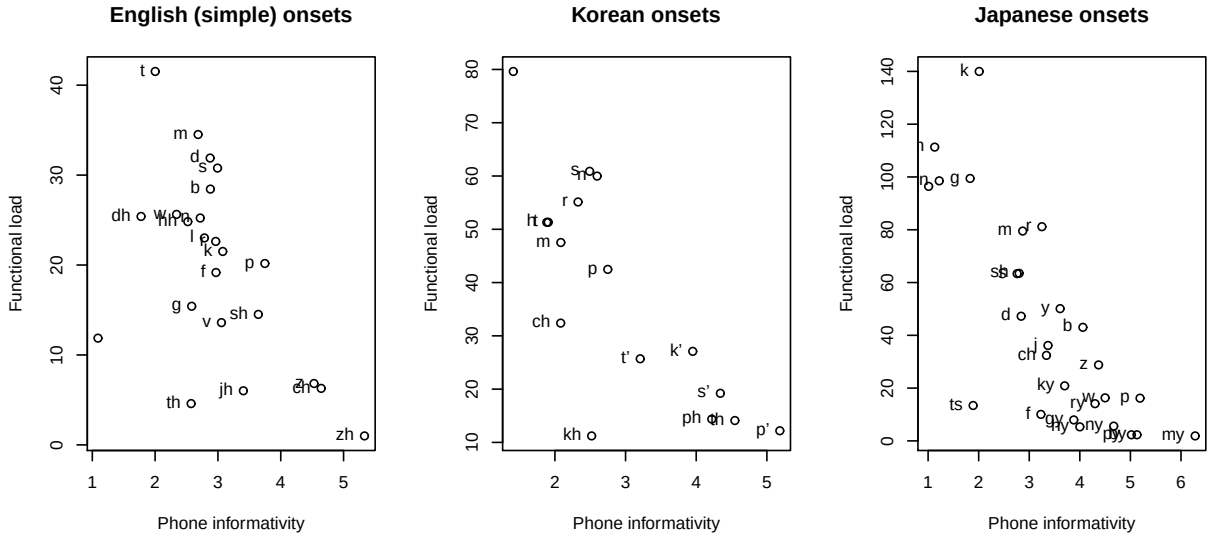

Figure 2.8: Cross-linguistic comparison of onset correlations.

2.6.2 Cross-linguistic comparison of onset and coda: Informativity

Figure 2.10 illustrates averaged PI of simple onsets and codas in English, Korean, and Japanese. In order to assess the effects of language (English, Korean, and Japanese), syllable position (onset, coda), and their interactions on the PI, the mean PI of onset and coda in each language was compared using Analysis of Variance. The result shows that there is a significant effect of language ($F(2, 98) = 4.134, p < 0.05^*$) and syllabic position ($F(1, 98) = 10.481, p < 0.01^{**}$), as well as their interaction ($F(2, 98) = 22.45, p < 0.01^{**}$). The post-hoc comparison using Tukey HSD method reveals that the mean PI of Japanese codas is significantly different from that of codas in other languages. Also, the mean PI of coda is significantly higher than that of onset in Japanese, which may be explained by the prevalence of codaless syllables in Japanese. In contrast, mean PI is comparable between onset and coda in English and Korean.

Also worth noting is that variances of PI distribution across the three languages are quite different, although the overall mean is similar. For example, the variability in English coda is larger than those in Korean and Japanese codas. Similarly, the variability in Japanese onsets is larger than those of onsets in English and Korean. This language-specificity is partly attributed to the difference in the number of attested phonemes in either onset or coda between the three languages.

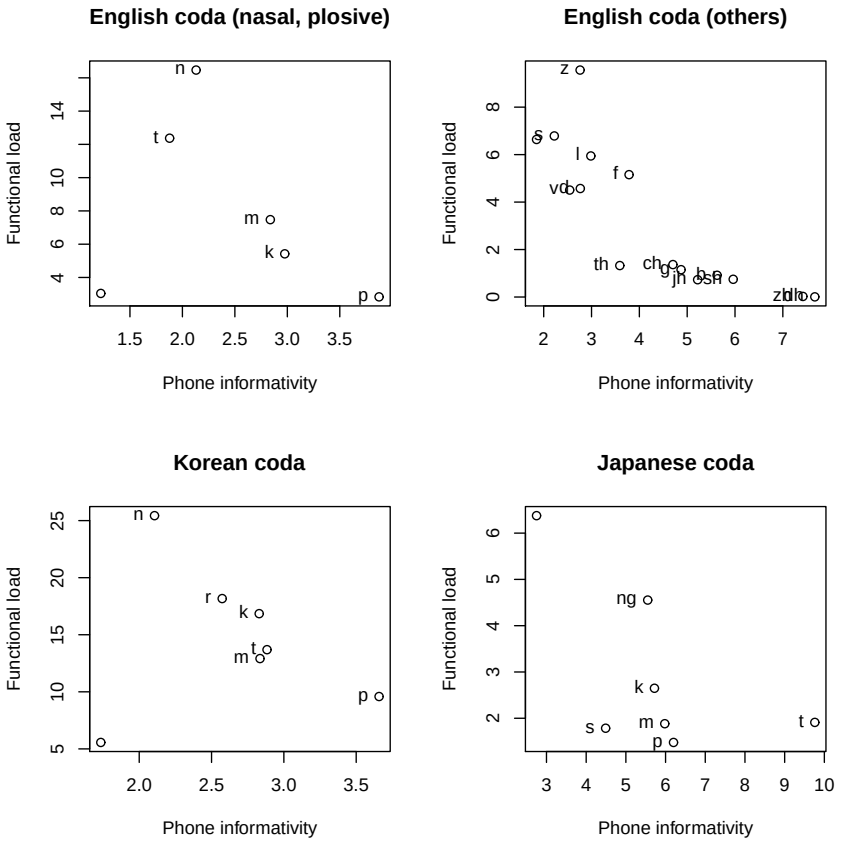

Figure 2.9: Cross-linguistic comparison of coda correlations. The results for English codas are depicted in two separate plots.

2.6.3 Cross-linguistic comparison of onset and coda: Functional load

A cross-linguistic comparison was carried out for onset FL and coda FL. The overall mean for English, Korean, and Japanese is shown in Figure 2.11. Because log scale better depicts the distribution of FL, log-transformed FL is used.

Similar to PI, the effects of language (English, Korean, and Japanese), syllable position (onset, coda), and their interactions on FL were tested by Analysis of Variance. The result shows that there is a significant effect of language ($F(2, 95) = 12.182, p < 0.01$*) and syllabic position ($F(1, 95) = 56.188, p < 0.01$**). There is also a marginal effect of their interaction ($F(2, 95) = 2.807, p = 0.06$) on FL. The post hoc comparison using Tukey HSD method shows that while the mean FL of onsets is not significantly different across the three languages,

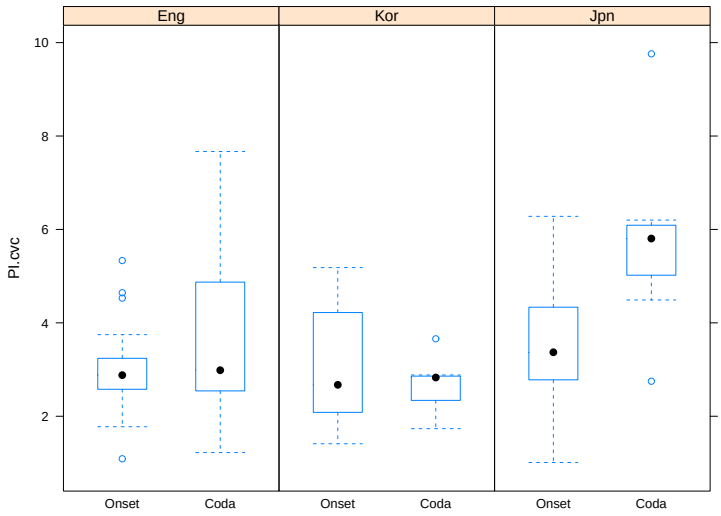

Figure 2.10: Averaged PI for simple syllable onset and coda in English (left), Korean (middle), and Japanese (right).

the mean FL of Korean codas is significantly higher than that of both English codas and Japanese codas at $p < 0.05$.

The result suggests that Korean codas, on average, convey a relatively larger amount of information than the codas in the other two languages. This observation may be explained by the fact that although syllable coda is common in Korean, a relatively strict restriction exists as to what could be coda. This is interesting because the two other languages show consistency in terms of the frequency of coda occurrence and the size of inventory of attested codas: Coda occurs frequently in English syllable and there is a relatively larger selection of phones that could occur in coda; but in Japanese, the occurrence of syllable coda is very limited, and even when it occurs, only a few phones can occur as coda.

2.6.4 Cross-linguistic comparison between CV units and VC units

Figure 2.12 shows the averaged FL of CV and VC sub-syllabic units for English, Korean, and Japanese. Compared to the FL of individual phones, the number of observations is much larger, resulting in a greater range for variability. The effects of language (English, Korean and Japanese), syllable position (onset and coda), and their interactions on FL were tested by Analysis of Variance.

A significant effect was found for language ($F(2, 2352) = 139.338, p < 0.01$), syllabic

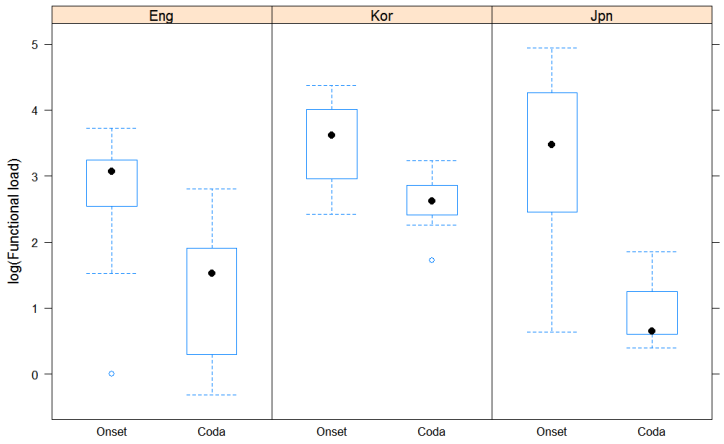

Figure 2.11: Averaged FL for syllable onset and coda for English (left), Korean (middle), and Japanese (right).

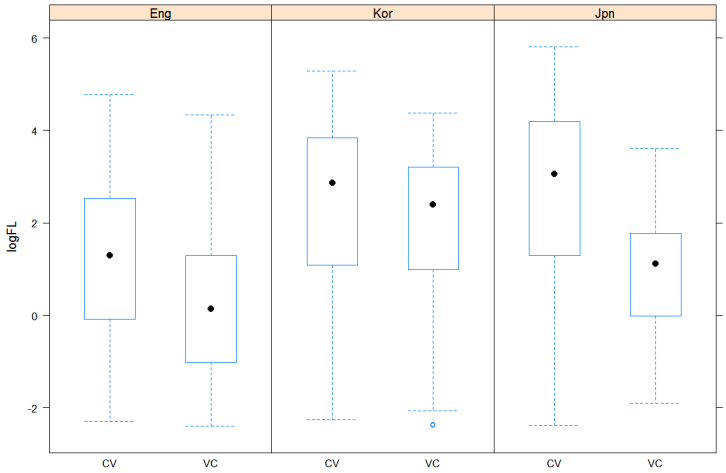

Figure 2.12: Averaged FL for sub-syllable CV and VC units in English (left), Korean (middle), and Japanese (right).

position ($F(1, 2352) = 301.972, p < 0.01$), and their interaction ($F(2, 95) = 7.001, p < 0.01$). A post-hoc analysis using Tukey HSD shows that the mean FL of English significantly differs from that of both Korean and Japanese at $p < 0.05$, but the mean FL of Japanese and Korean were not different from each other. In addition, the post-hoc test also shows that the mean FL of CV is significantly greater that that of VC at $p < 0.05$. However, the mean FL between onset and coda in Korean were not significantly different from each other; in contrast, the

FL of CV units were significantly higher than that of VC units in both English and Japanese ($p < 0.05$). The result is consistent with the findings from the mean FL of the individual phones, where codas conveyed a relatively higher FL in Korean.

This comparison between Korean and English is surprising given the previous finding on a greater degree of the association of CV unit than that of VC unit in Korean (Lee & Goldrick 2011). If the degree of association is positively related to the amount of information, one might expect that coda or VC unit in English would contain a greater amount of information than the codas in Korean. However, the pattern found in the FL reports the opposite: VC units or codas carry greater functional load in the sub-lexical σ-universe in Korean.

Chapter 3

Experiment 1: Perceptual Accuracy of Audio-only Stimuli

3.1 Chapter introduction

The current chapter investigates how the perceptual accuracy of listeners with different language backgrounds is related to the information measures computed in Chapter 2. Chapter 2 showed that the syllable structure of a language affects the distribution of the amount of information transmitted through different speech units, which is specific to each language. This chapter tests if the amount of information carried by individual speech elements correlates with their perceptual accuracy.

The chapter finds that two distinct information measures, PI and FL, computed under different specifications, show varying correlations with perceptual accuracy, with apparent language-specificity. Later in this chapter, this language-specificity is interpreted as reflecting the characteristics of each language, such as its syllable structure.

3.2 Background of the experiment

In order to better understand the effects of information measures on perceptual accuracy, an identification paradigm for non-words is used, which is similar to the one used by Winters (2001). By using non-words, the experiment prevent lexical factors, which are not relevant for the current study, from affecting the listeners. Also, stop consonant clusters (CC) are used in the experiment because the three languages studied, English, Japanese and Korean, all have phonemic /p, t, k/ in onset and at least their phonetic realization in coda.

The use of stop consonant clusters is also related to the amassed findings in the assimilation literature. Substantial evidence has found asymmetries in assimilation patterns cross-linguistically (e.g. Jun 2004), and many researchers suggested that the perceptibility, or perceptual salience, is the underlying cause for such asymmetry. In this view, the perceptually weak sounds are likely to neutralize and undergo assimilation, often termed as

a	/mit+ko/	[mikk'o]	'believe and'
	/mit+pota/	[mipp'ota]	'more than the bottom'
b	/ip+ko/	[ikk'o]	'wear and'
	/nop+ta/	[nopt'a] *[nott'a]	'be high'
c	/nok+ta/	[nokt'a] *[nott'a]	'melt'
	/kuk+pota/	[kukp'ota] *[kupp'ota]	'more than soup'

Table 3.1: Place assimilation in Korean (adapted from Jun 1996)

"perceptually tolerated articulatory simplification" (Kohler 1990, Hura et al. 1992, Huang & Johnson 2008). According to this account, the difference in the ways English, Korean, and Japanese realize consonant clusters may cause the differential assimilation pattern as well. For instance, in Japanese, stop codas alone do not occur in a syllable; however, stop codas always fully assimilate to the following sound in terms of its place. In Korean, syllable codas are allowed, but the attested set of codas is much more limited than the set of onsets due to the neutralization rule (shown in (2.5) in Chapter 2). Lee and Goldrick (2011) also mentioned this when explaining the stronger association of the body(CV-) unit in Korean that most of the weakly associated codas neutralize to be one of /p, t, k/, causing the three stops to be much more frequent in the surface form than they are in the underlying forms. Interestingly, even when the two stop CCs are attested in Korean, stop codas are not always fully realized, causing the asymmetric pattern of place assimilation in production (Jun 1996). Coronal /t/, when placed before other stop sounds, is likely to assimilate to the following sounds, labials assimilate only before velars, and finally, velars never assimilate (as shown in Table 3.1). Similarly, in an actual production pattern, coronals are often deleted, and labials are often produced with more gestural overlap than velars in syllable coda before another stop sound (Kochetov, Pouplier & Son 2007). Finally, English, unlike Japanese or Korean, has a relatively flexible set of stop-stop sequences with a smaller degree of place assimilation.

To summarize, the specific research questions addressed by this chapter are:

Question 1 Can perceptual accuracy be predicted from the information measures computed from the syllable universe? If so, how do information measures correlate with perceptual accuracy?

Question 2 Do speech elements of a certain size better predict perceptual accuracy than other types of speech elements, for listeners of each of the three languages? To answer this question, FL computed for phonemes[1] is compared with that computed for CV

[1]In this chapter, phonemes are used interchangeably with phones in describing experiment stimuli; the stimuli consist of stop sounds /p/, /t/, and /k/, which are distinct phonemes in all the three languages.

and VC units. For PI, measures computed under different definitions of contexts are compared.

Question 3 Do the perceptual patterns of English, Japanese and Korean listeners differ by syllabic position (onset vs. coda)?

3.3 Method

3.3.1 Stimuli

The experiment uses sequences consisting of two stops, VC.CV, with '.' denoting the syllable boundary.

V	C.C	V
	p.p	
	p.t	
	p.k	
a	t.p	a
u	t.t	u
i	t.k	i
	k.p	
	k.t	
	k.k	

Table 3.2: Recording syllable list

A comprehensive set of VC and CV syllables was recorded, with one of the vowels /a, u, i/ and one of the consonants /p, t, k/, as shown in Table 3.2.

Recording The current experiment used naturally spoken recording with little acoustic editing to create natural-sounding stimuli. Eight different talkers were recruited for stimulus recording. Four were native American English talkers, and the other four were native Korean talkers. Two from each group were male and the other two were female. Each of them was a dominantly native talker of the target language, and had lived in an environment where the target language was spoken dominantly at least until the age of 19.

Before the recording session, the talkers were informed that each VC or CV syllable was a non-word and each token would appear on the screen one at a time, during which they would repeat the syllable three times at a tone that was as flat as possible. English talkers were specifically told that each word they would see on the monitor was written in IPA, rather than in real alphabet, such that a syllable like /ap/ should be read as [ap] instead of /æp/. All of the four English talkers were familiar with IPA symbols. The non-words

for Koreans were also written in a phonetically specific alphabet, the Korean orthography (Hangul).

After instruction and training, the talker sat in a sound-attenuated room in front of a computer monitor with a standing AKG 535 EB microphone placed in front of him or her. Each token appeared on the screen in a random order on a PowerPoint slideshow. The talker read and produced each syllable three times consecutively. When the syllable was misspoken, spoken with a wrong intonation, or spoken with a different pronunciation from the intended one, the lab attendant sitting next to the talker corrected the talker. The productions were digitally recorded by a separate PC inside the sound booth at a 44K sampling rate using Praat (Boersma & Weenink 2007). The faces of the talkers were also videotaped for an audio-visual experiment (using Canon Model XF 100a), which will be described by Chapter 4.

Editing The recording was carefully inspected using a waveform editor, Praat (Boersma & Weenink 2007), by two trained phoneticians. For each syllable, the token with the least fluctuation in intonation was selected as the base syllable. For VC syllables spoken by the English talkers, the release burst at the end of a stop was removed to make the stimuli comparable to those by Korean talkers.[2]

In the end, 18 syllables, six syllables for each vowel (e.g. /ap/, /at/, /ak/, /pa/, /ta/, /ka/), were selected for each talker. The base syllables selected for each talker were first normalized to equal average amplitude using SoX[3], then spliced back-to-back to create non-overlapping VC.CV-stimuli. All the possible combinations of consonant clusters were created with the same vowel used twice within a VC.CV token (e.g. [at.ka] was created, while [ut.ka] was not). In total, 216 stimuli (9 consonant clusters x 8 talkers x 3 vowels) were created.

3.3.2 Participants

A total of 55 listeners participated in a nine-alternative, forced-choice identification experiment.

Twenty English listeners were recruited for the experiment. All of the participants were attending University of California at Berkeley at the time of the experiment. None of them reported a speech hearing disorder or problem. Fifteen native Korean listeners were recruited

[2]It is well-known that release bursts of post-vocalic (coda) stops affects and changes perceptibility of stops (Kochetov & So 2007, Malécot 1956, Repp 1984). Although the production of English stops often accompanies a release burst (Henderson & Repp 1982), or at least is considered a more canonical form among many allophonic variants (Sumner & Samuel 2005, Chang 2014), data from speech corpora suggest that a substantial proportion of stops are produced unreleased (Davidson 2011). Furthermore, in a context such as consonant clusters, the rate of the occurrence of unreleased stop clusters is even greater intramorphemically and intermorphemically (Bergier 2014). In light of these findings, the release burst in the current study was removed under the assumption that this controls for factors that may affect the perception without making the stimuli overly artificial.

[3]SoX is a free digital audio editor, licensed under the GNU General Public License

from Seoul National University in Seoul, South Korea. They all had no or very little experience living in an English-speaking country and had started learning English after the age of 10. Sixteen Japanese listeners were recruited from Daito Bunkyo University in Tokyo, Japan.

The number of participants from the three language groups is summarized in Table 3.3.

Language	Male	Female	Total
English	8	12	20
Korean	8	7	15
Japanese	5	11	16

Table 3.3: Summary of the participants from the three language groups. Male and Female columns refer to the number of participants in each group.

3.3.3 Procedure

The speech perception task was a nine-alternative forced-choice identification (e.g. Hura et al. 1992 and Winters 2003). The specific experimental environments were different for each language group since the experiment took place in different places for each group. However, the overall procedure for the experiment was identical across the language groups.

The participants were first given instructions for the experiment, then were seated in a sound-attenuated room for the perception experiment. The experiment stimuli were then presented using Opensesame software (Mathôt et al. 2012). Participants wore AKG K 240 headphones and were given a volume-adjusting switch so that they could freely adjust the volume to their comfort level.

Participants listened to one of the randomly presented 216 stimuli and identified the intervocalic consonants as one of the nine alternatives ("pt", "pk", "pt", etc.). The nine choices appeared on the screen in a 24-point font as a vertical list, and the participants selected one of them by clicking on it using a mouse attached to a PC. The next token was presented as soon as a participant selected one of the choices. The entire set of 216 was presented twice, in two different random orders.[4]

[4]Originally, the experiment was carried out in two different conditions: normal condition vs. attention-manipulated condition. Under the attention-manipulated condition, the attention of a participant was divided by the additional task of keeping the count of tokens given so far. The rationale for the two different conditions was that in the perception of a non-word, listeners may not show any language-specificity under a normal condition. Indeed, recent production literature has shown that talkers do not clearly reveal their bias unless they are explicitly 'required' to engage in such bias. In other words, as long as there is sufficient cognitive capacity, any underlying bias is not shown. The difficulty of capturing the bias in low-level tasks is also shown in the literature on attention. Hon and Tan (2013) concluded that if the attention is the key factor that leads to certain differential patterns between two conditions, manipulating the attention directly would magnify the pattern found in the regular condition. They also found that having a 'harder' task would

In total, each participant listened to 432 tokens (8 talkers * 3 vowels * 3 codas * 3 onsets * 2 iterations).

3.3.4 Analysis

3.3.4.1 Information measure 1: Informativity

To compute the information measures of individual consonants in coda and onset (coda is the first consonant in a VC.CV sequence and onset is the second consonant), a corpus of VC.CV sequences is constructed. For both coda and onset, the adjacent vowel or the adjacent consonant is used as the context in computing PI.

PI measures computed and used in this chapter are distinct from those used in Chapter 2, which were computed from a syllable corpus. The corpus of VC.CV sequences is adopted in this chapter because the form of speech elements in the corpus is identical to that of the experimental stimuli. Such a corpus would preclude any potential bias that may arise due to a mismatch between the form of corpus elements and that of experimental stimuli.

Type	Context (coda)	Context (onset)
PI.V	(a)p.ka.	ap.k(a)
PI.C	ap.(k)a	a(p).ka

Note: The relevant context in each case is enclosed by ().

Table 3.4: Example of two different ways of how PI can be defined for a non-word /ap.ka/.

For notational convenience, the PI of coda and onset with the adjacent vowel as the context is denoted by PI.V, and the PI with the adjacent consonant as the context is denoted by PI.C. Table 3.4 illustrates this notation with a non-word /ap.ka/ as an example.

3.3.4.2 Information measure 2: Functional load

Unlike PI, the definition of FL does not depend on any nearby sound. FL is computed for each onset, coda, VC unit and CV unit using the syllable corpus presented in Chapter 2.

3.3.4.3 Statistical analysis

Listeners' responses from the perception experiment were analyzed through logistic mixed-effect regression modeling (Baayen 2008, Jaeger 2008). The models were fit with the R package lmer (Bates 2014).

remove the ceiling effect from the regular condition. Indeed, the pilot data showed that listeners, regardless of language background, were fairly good in all stimuli. This ceiling effect is partly attributable to the nature of the experiment, which does not require much cognitive load. Also, the attention manipulation resulted in no significant effects in a preliminary statistical analysis; therefore, it is ignored in the main analysis.

The dependent variable was a listener's response to a VC.CV token coded as either 1 (correct) or 0 (incorrect), which was fitted with broadly three types of fixed factors: (1) stimulus factors; (2) a listener factor; and (3) information measures, which were critical.

Stimulus factors include various linguistic and physical aspects of the sound stimuli that may have affected listeners' perception, regardless of their language background. These include phones of interest (p, t, k), adjacent consonants (p, t, k), adjacent vowels (a, u, i), syllable position (coda vs. onset), and talker language (Korean vs. English). The eight different talkers were coded to identify their native language only because including their identity as a fixed factor would increase the number of model parameters too much.

The listener factor, which codes a listener's language (ListenerLang in the model), is included to capture the cross-linguistic difference in the degree to which information measures affect perception. More specifically, the interaction terms between the listener factor and information measures are used to capture this difference.

For PI, as explained earlier, two different definitions of a context are used; PI.V is computed with the adjacent vowel as the context and PI.C is computed with the adjacent consonant as the context. Also, two types of FL are computed, which differ by the size of the speech element for which FL is computed: FLseg is the FL of an individual consonant (segmental level), while FLsub is the FL of a VC or a CV unit (sub-syllabic level). For PI, as explained earlier, two different definitions of a context are used; PI.V is computed with the adjacent vowel as the context and PI.C is computed with the adjacent consonant as the context. Also, two types of FL are computed, which differ by the size of the speech element for which FL is computed: FLseg is the FL of an individual consonant (segmental level), while FLsub is the FL of a VC or a CV unit (sub-syllabic level).

For each of these four types of information measures, a model is built with as many relevant fixed factors as possible. Some of the factors are dropped because they do not have any explanatory power, and the selection of factors, or the model selection, is done through χ^2 likelihood ratio tests and comparisons of AIC and BIC.

For each language and for each of the different information measures, the information measure is linearly normalized so that the minimum has the value of 0 and the maximum has the value of 1. This normalization makes it easier to compare results across different languages and information measures. Since there are only several (three or nine) observations for each language, this normalization is more convenient for displaying results than more common z-score normalization is; z-scores can produce wildly varying ranges of values across different languages and information measures. Moreover, this normalization produces the same predicted accuracy as z-score normalization does, since they are both linear functions of original information measures.

In addition, the random effects (random slopes) at the level of subjects and talkers are included. It is done to minimize type-I error by creating a maximal random effect model. A more detailed description of the model structure is discussed in the following sections.

Whenever a full model including all the factors failed to converge, the set of factors was reduced by excluding factors, one at a time, starting with the one that makes the least contribution to model-fit, which was measured by the χ^2 likelihood ratio test.

The coding convention for the factors used in describing the models is as follows:

- Fixed Effects:
  - Stimuli factors
    - SOUND_LABEL: /p, t, k/. Physical or auditory characteristics of /p, t, k/ inherent in the sound.
    - SYLL.POS: The position of a phoneme in a syllable, which is onset or coda.
    - NEARC: /p, t, k/. The adjacent consonant in the VC.CV sequence.
    - VOWEL: /a, u, i/. The adjacent vowel in the VC.CV sequence.
    - TALKERLANG: The language of a talker, which is Korean or English. A listener with the same language background may have a perceptual advantage due to language familiarity.
  - Listener factors
    - LISTENERLANG: The native language of a listener, which is English, Japanese or Korean. It would presumably affect the degree to different information measures predict perceptual accuracy
  - Information factors
    - PI.V: Phonological informativity computed from the VC.CV sequences with the adjacent vowel as the context.
    - PI.C: Phonological informativity computed from the VC.CV sequences with the adjacent consonant as the context.
    - FL.seg: Functional load computed for individual consonants.
    - FL.sub: Functional load computed for VC and CV units.
- Random Effects:
  - SUBJECT: A unique subject number given to each participant.
  - TALKER_ID: A unique talker number given to each talker.

3.4 Result and discussion

3.4.1 Informativity

3.4.1.1 Overview

The results in this section show how listeners' perceptual accuracy changes as a function of the two types of phonological informativity (PI). A model containing all the observations is constructed with as many potentially relevant controlling and critical factors as possible, but some of the factors that do not improve the model fit are excluded. Since two different

PI measures, PI.C and PI.V are available, a separate model is constructed with each of the two measures.

The two models serve multiple purposes. To begin with, they determine the signs of the effects of PI. They also reveal the cross-linguistic difference, through the interaction terms between PI and a listener's language. Finally, by comparing the two models, one based on PI.C and the other on PI.V, it is possible to observe the cross-linguistic difference as to which of the two measures more strongly correlates with perceptual accuracy.

In addition, smaller models specific to each syllable position were constructed to analyze onset and coda separately (position specificity is motivated by Raymond et al. (2006) and Cohen Priva (2008)). However, some of those models failed to converge, and others showed inconsistent results. Therefore, they are not discussed in this chapter.

3.4.1.2 Model estimation

Model structure The baseline (or 'null') model is constructed with multiple controlling factors including stimuli-inherent factors and listener factors. In capturing the language-specific effect of PI, the critical factor of interest is the interaction term between PI and a listener's language ListenerLang. As for random effects, random intercepts for each subject and talker (written as talker_id for coding convention) as well as a by-subject random slope for listener language are included.

With the fixed and random factors described above, the probability of a correct response in the comprehensive model is stated by the following formula:

$$Prob(Correct = 1) = P[\alpha_1 + \alpha_2 + \sum_x \beta_x PI \cdot I(ListenerLang = x) + \mathbf{\Gamma} \cdot (\text{Control Variables})]. \tag{3.1}$$

The equation in 3.1 represents a mixed effects logistic regression. P is the cumulative distribution function of a logistic distribution. α_1 and α_2 are subject and talker random effects, respectively. $\mathbf{\Gamma}$ is the vector of coefficients on the control variables, which are ListenerLang, Syll.pos, Sound_label, Vowel, NearC, and talker language-ListenerLang interaction. The coefficient β_x represents the effects of PI by listener language group, which is the main coefficient of interest. The variable x represents the three languages: English, Japanese and Korean.

The three-way interaction between PI, ListerLang and Syll.pos is not included because the logistic regression fails to converge with the three-way interaction term.[5] As an alternative, the effect of a syllabic position is studied by estimating separate models for onset and coda, which is discussed later.

The comprehensive model is estimated with each of the two PI measures, PI.C and PI.V, which yields two distinct set of coefficient estimates. 3.5 shows the estimated coefficients.

[5]It may be due to the insufficient number of observations for the given set of factors.

Effects of Controlling factors With either of the two PI measures, the controlling factors show similar effects. A positive coefficient on a factor indicates a positive relationship between that factor and perceptual accuracy. The effects of the controlling factors are graphically illustrated in Figures 3.1 and 3.2.

One of the most prominent effects is found in SYLL.POS and the perceptual accuracy for onset is much greater than that for coda. The acoustic properties of a sound also affect perception: Everything else being equal, /k/ shows the lowest accuracy, while /t/ elicits the highest accuracy. The adjacent vowel also affect perception: In both models, listeners' accuracy is lower for /i/ than for /a/ and /u/. The effects of the adjacent consonant are not as clear as those of other factors, but a χ^2 likelihood ratio test reveals that the factor still slightly improves the model fit ($\chi^2 = 5.961, df = 2, p = 0.051$), with /k/ indicating the lowest and /t/ indicating the highest probability of a correct answer.

There is also an effect from a listener' native language, as well as the interaction between a listener's and a talker's languages. Japanese listeners clearly exhibit overall lower accuracy than English and Korean listeners. In addition, compared to English listeners' responses to English talkers, their responses to Korean talkers show significantly lower accuracy in both PI.C and PI.V models. Such an advantage may be derived from the match between a listener's and a talker's languages, but it is not found for Korean listeners.

Effects of Critical factors The result indicates clear effects of PI and listener language. As shown in Figures 3.3 and 3.4, the interaction terms between PI and listener language groups indicate significant language-specific effects of PI. Korean listeners' response accuracy decreases as PI increases, but the extent to which it decreases is greater in the PI.C-model (β=-0.74, $p < 0.001$) than in the PI.V-model (β=-0.44, $p < 0.001$). In contrast, although Japanese listeners' response accuracy also decreases as PI increases, the slope is steeper in the PI.V-model (β=-0.62, $p < 0.001$) than in the PI.C-model (β=-0.39, $p < 0.001$).

Table 3.5: Estimated coefficients from the PI.C- (left) and PI.V-model (right).

	PI.V-model	PI.C-model
ListenerLang, Eng.L	0.200	-0.531^{***}
	(0.146)	(0.154)
ListenerLang, Jpn.L	-0.407^{**}	-0.568^{***}
	(0.170)	(0.169)
Syll.pos, Onset	1.957^{***}	2.223^{***}
	(0.039)	(0.037)
Sound_label, t	0.275^{***}	0.332^{***}
	(0.034)	(0.035)
Sound_label, k	-0.520^{***}	-0.488^{***}
	(0.034)	(0.039)
Vowel, i	-0.193^{***}	-0.195^{***}
	(0.030)	(0.030)
Vowel, u	0.005	0.005
	(0.030)	(0.030)
NearC, p	0.061^{**}	0.062^{**}
	(0.030)	(0.030)
NearC, t	0.065^{**}	0.065^{**}
	(0.030)	(0.030)
Kor.L * Kor.T	-0.040	-0.040
	(0.233)	(0.233)
Eng.L * Kor.T	-0.396	-0.389
	(0.266)	(0.260)
Jpn.L * Kor.T	-0.186	-0.194
	(0.132)	(0.137)
Kor.L * PI	-0.440^{***}	-0.733^{***}
	(0.065)	(0.098)
Eng.L * PI	-0.244^{***}	0.621^{***}
	(0.076)	(0.084)
Jpn.L * PI	-0.603^{***}	-0.383^{***}
	(0.085)	(0.059)
Constant	0.747^{***}	0.669^{***}
	(0.185)	(0.185)
Observations	42,336	42,336
Log Likelihood	$-20{,}549.230$	$-20{,}506.660$
AIC	41,144.460	41,059.320
BIC	41,343.490	41,258.350

Note: Standard errors of the coefficients are given in parentheses.
Coefficients that are significantly different from zero are marked with asterisks.
($^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$)

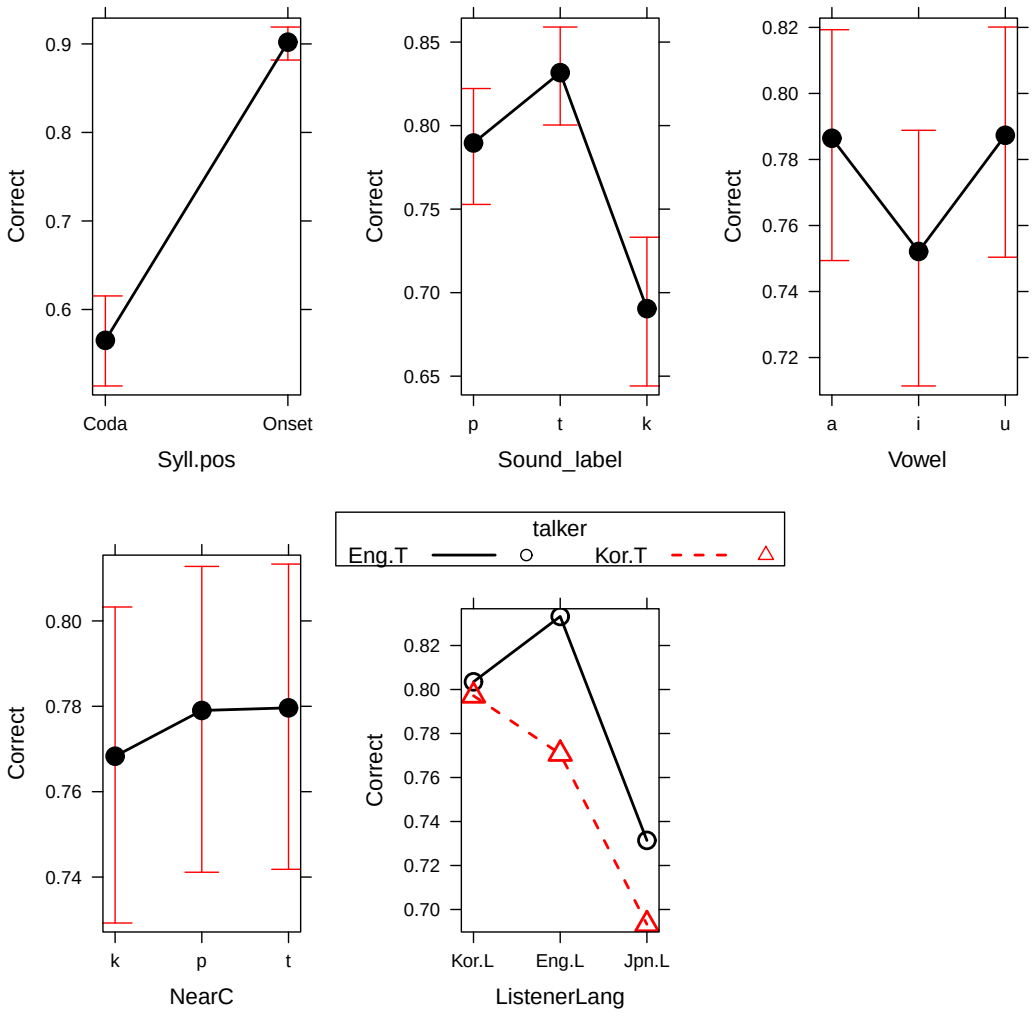

Figure 3.1: Partial effect plots for controlling fixed factors for the PI.V-model: Syllable position, adjacent vowel and consonant, and sound.

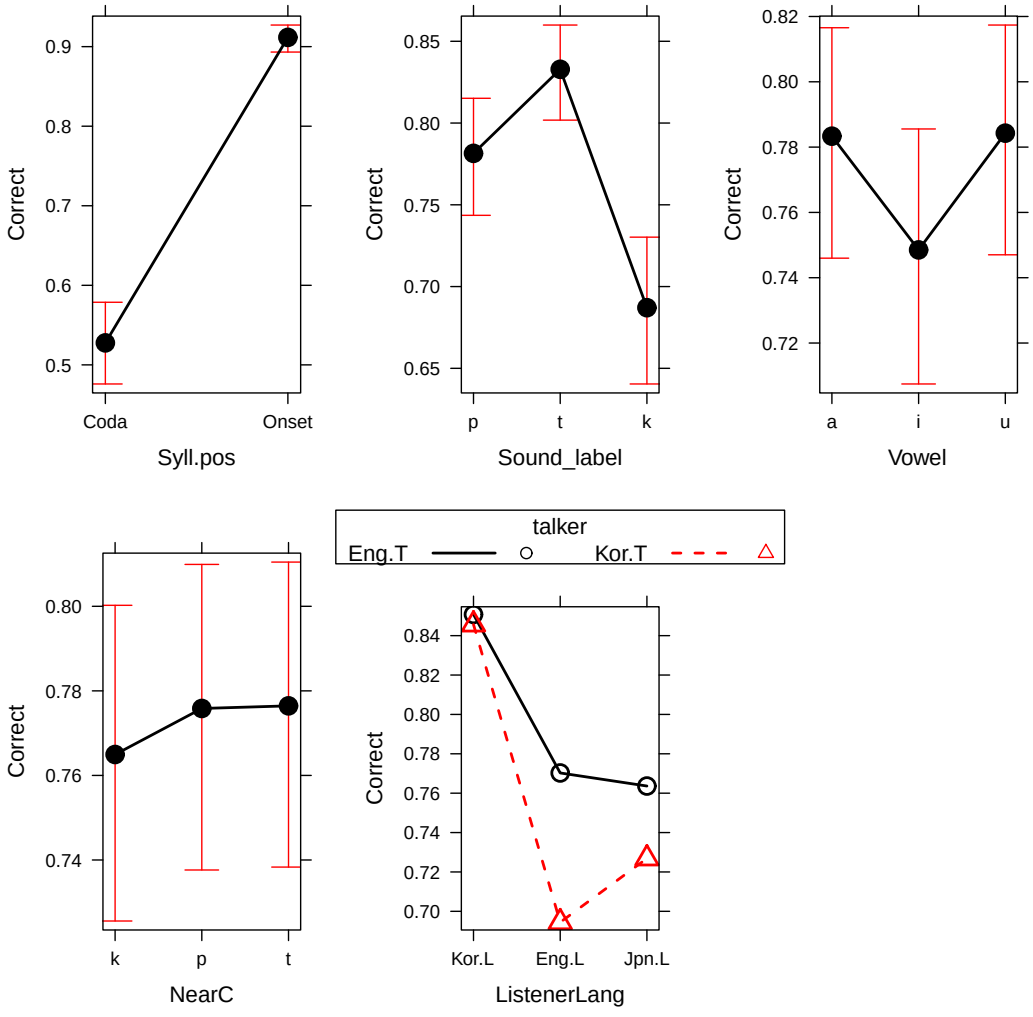

Figure 3.2: Partial effect plots for fixed factors for the PI.C-model: Syllable position, adjacent vowel and consonant, and sound.

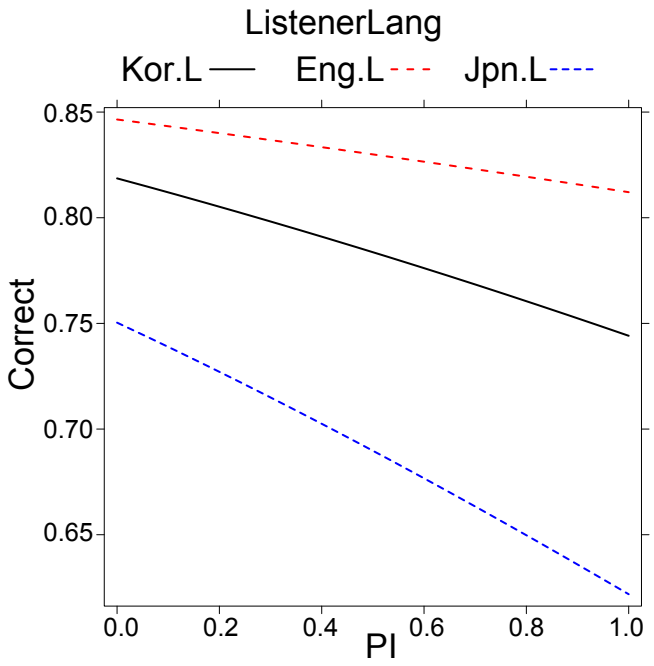

Figure 3.3: Model predictions for the interaction of PI.V and listener language.

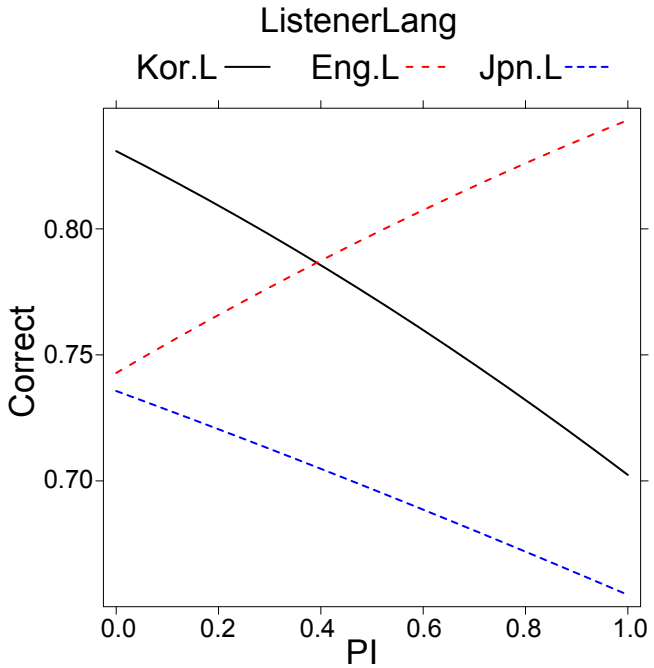

Figure 3.4: Model predictions for the interaction of PI.C and listener language.

3.4.1.3 Discussion

Except for the case of English listeners with PI.C, PI correlates negatively with perceptual accuracy. This result implies that 'contextual familiarity' generally has a positive effect on perceptual accuracy; the contextual familiarity is defined as the negative of PI, and represents familiarity in the sense that it is increasing in the probability that the phoneme of interest occurs under relevant contexts.

The observation that PI.V shows a stronger correlation with perceptual accuracy than PI.C for Japanese listeners may be explained by the fact that CV syllables are prevalent in Japanese; Yoneyama (2000) showed that CV syllables occupy over 90 percent of the entire set of Japanese syllables. With such a prevalence of CV syllables, a consonant often would not be adjacent to any other consonant, and listeners would have relatively little experience in processing a consonant cluster. For example, in a sequence of CV syllables only, no pair of consonants are adjacent to each other.

In addition, the different relative effects of PI.C and PI.V between Japanese and Korean may indicate different perceptual units important in each language. The greater sensitivity to PI.C, exhibited by Korean listeners, may indicate that Korean listeners are relatively used to processing multiple syllables as a perceptual unit, as opposed to a single syllable as a perceptual unit. In both Korean and Japanese, a consonant and its adjacent consonant are always in adjacent syllables, but never in the same syllable.

3.4.2 Functional load

3.4.2.1 Overview

This section studies perceptual accuracy as a function of the two types of functional load (FL): FL.seg, which is computed for individual phonemes, and FL.sub, which is computed for VC and CV units. This is different from PI-based models, for which the two types of PI measures are distinguished by the definition of a context.

The method of constructing models is similar to that used in the previous subsection. The following controlling factors are included: target consonant (SOUND_LABEL; for FL.sub, it is the consonant within the VC or CV unit), neighboring vowel VOWEL; for FL.sub, it is the vowel within the VC or CV unit), neighboring consonant (NEARC), and talker language (TALKERLANG). The dependent variable is the binary response from the listeners: Correct vs. incorrect, coded as 1 and 0, respectively. SUBJECT and TALKER_ID are taken as the random factors. Finally, FL measures are linearly rescaled to range between 0 and 1 for each language in order to compare the three languages more easily.

As is done with PI, two comprehensive models with as many relevant fixed factors as possible is first constructed with FI.seg and FL.sub each. In addition, with FL.seg, an onset-specific model is constructed, using data points only for onset. Similarly, with FL.sub, an onset-specific model is constructed, using data points only for VC units. Coda-specific models were also tried but they failed to converge, and thus are not discussed in this chapter.

3.4.2.2 Comprehensive models

Model structure Similarly with PI, mixed effect logistic regression models are used:

$$Prob(Correct = 1) = P[\alpha_1 + \alpha_2 + \sum_x \beta_x FL \cdot I(ListenerLang = x) + \mathbf{\Gamma} \cdot (\text{Control Variables})]. \tag{3.2}$$

P is the cumulative distribution function of a logistic distribution. α_1 and α_2 are subject and talker random effects, respectively, and $\mathbf{\Gamma}$ is the vector of coefficients on the control variables, which are ListenerLang, Syll.pos, Sound_label, Vowel, NearC, and the interaction of talker language with ListenerLang. The coefficient β_x represents the effects of FL (either FL_{SEG} or FL_{SUB}) by listener language group, which is the main coefficient of interest. The variable x represents the three languages: English, Korean, and Japanese.

For the FL_{SEG}-based model, no random slope for listener language group is included, as it does not improve the model fit. For the FL_{SUB}-based model, a by-talker random slope for listener language group is included, as it improves the model fit significantly.

Table 3.6 shows the estimated coefficients.

Controlling factors The effects of the controlling factors are similar to what is found with PI, and therefore not discussed further.

Critical factors The effects of the critical factors are very similar between FL.seg and FL.sub. As Figures 3.5 and 3.6 illustrate, both FL.seg and FL.sub have a positive effect for English and Japanese listeners, while they have have a slightly negative effect for Korean listeners.

Also, the degree to which FL.seg affects perceptual accuracy is significantly greater for Japanese listeners (β=0.75, p <0.001***) than for English listeners (β=0.40, p <0.001***). FL.sub also has a greater effect on the perceptual accuracy for Japanese listeners than for English listeners.

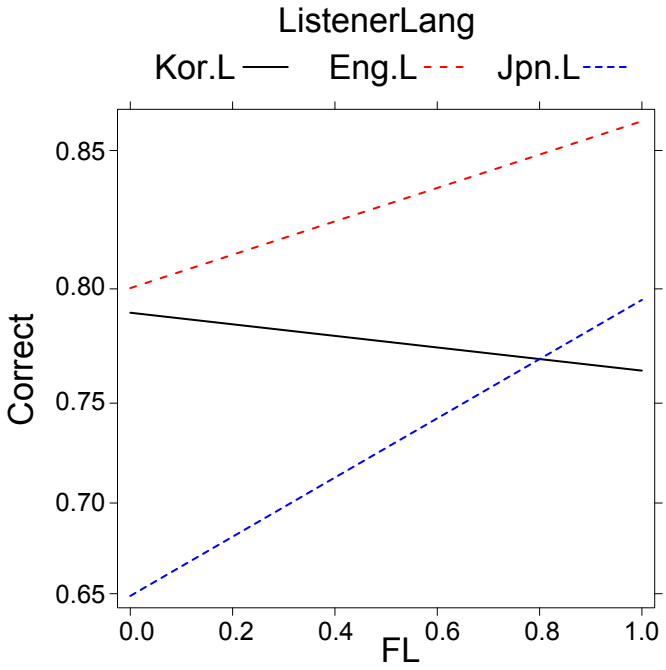

Figure 3.5: Model predictions for the interaction of FL.seg and listener language.

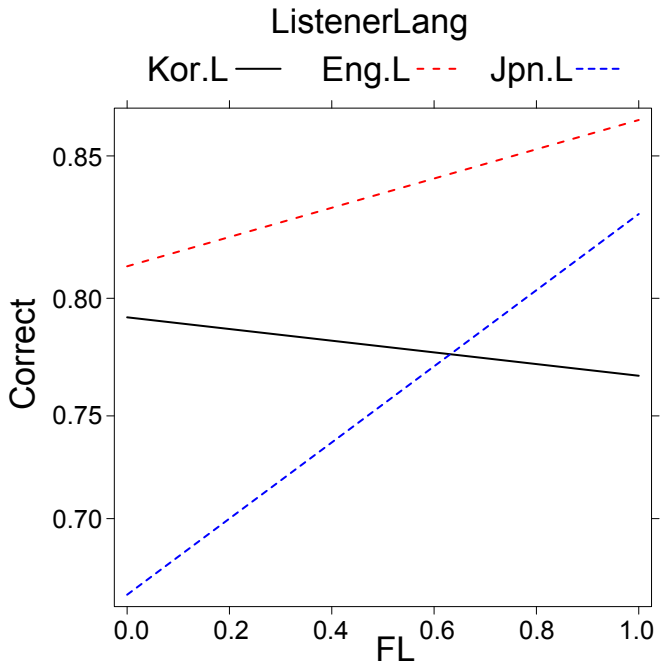

Figure 3.6: Model predictions for the interaction of FL.sub and listener language.

Table 3.6: Estimated coefficients from the FL.seg- (left) and FL.sub-model (right).

	FL.seg	FL.sub
ListenerLang, Eng.L	0.062	0.125
	(0.134)	(0.134)
ListenerLang, Jpn.L	-0.713^{***}	-0.679^{***}
	(0.164)	(0.166)
Syll.pos, Onset	1.863^{***}	1.957^{***}
	(0.040)	(0.032)
Sound_label, t	0.140^{***}	0.197^{***}
	(0.035)	(0.031)
Sound_label, k	-0.486^{***}	-0.488^{***}
	(0.032)	(0.030)
Vowel, i	-0.194^{***}	-0.169^{***}
	(0.030)	(0.031)
Vowel, u	0.005	0.024
	(0.030)	(0.033)
NearC, p	0.061^{**}	0.061^{**}
	(0.030)	(0.030)
NearC, t	0.065^{**}	0.065^{**}
	(0.030)	(0.030)
Kor.L * Kor.T	-0.041	-0.041
	(0.233)	(0.236)
Eng.L * Kor.T	-0.393	-0.392
	(0.264)	(0.262)
Jpn.L * Kor.T	-0.189	-0.189
	(0.134)	(0.134)
Kor.L * FL.seg	-0.146^{**}	
	(0.069)	
Eng.L * FL.seg	0.420^{***}	
	(0.102)	
Jpn.L * FL.seg	0.745^{***}	
	(0.076)	
Kor.L * FL.sub		-0.143
		(0.089)
Eng.L * FL.sub		0.358^{***}
		(0.128)
Jpn.L * FL.sub		0.931^{***}
		(0.096)
Constant	0.647^{***}	0.580^{***}
	(0.184)	(0.188)
Observations	42,336	42,336
Log Likelihood	$-20{,}543.720$	$-20{,}544.290$
AIC	41,133.450	41,134.580
BIC	41,332.480	41,333.610

Note: Standard errors of the coefficients are given in parentheses.
Coefficients that are significantly different from zero are marked with asterisks.
($^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$)

3.4.2.3 Onset-specific model

To further analyze the pattern observed from the comprehensive models, the effects of the two FL measures are estimated only for onset (the second consonant in VC.CV). As before, the responses are studied by two models, each using one of the two FL measures, FL.seg and FL.sub. The models are identical to the comprehensive models, except that some of the factors are no longer used. The factor SYLL.POS is no longer relevant, because only observations (responses) on onsets are used. The factor NEARC is removed to ensure model convergence and a by-talker random slope for a listener's language is included. The resulting expression for the models is very close to the one for the comprehensive models:

$$Prob(Correct = 1) = P[\alpha_1 + \alpha_2 + \sum_x \beta_x FL \cdot I(ListenerLang = x) + \mathbf{\Gamma} \cdot (\text{Control Variables})]. \tag{3.3}$$

P is the cumulative distribution function of a logistic distribution. α_1 and α_2 are subject and talker random effects, respectively, and $\mathbf{\Gamma}$ is the vector of coefficients on the control variables, which are LISTENERLANG, SOUND_LABEL, VOWEL, and the interaction of talker language with LISTENERLANG.

The results from the two models, one with FL.seg and the other with FL.sub, are shown in Table 3.7. The controlling factors generally elicit similar patterns to those found in the comprehensive models. The result indicates a clear language-specific effect of FL.seg and FL.sub. Also, the likelihood test confirms the significant effects of the critical factors (FL.seg: $\chi^2 = 99.915, df = 3, p < 0.001$; FL.sub: $\chi^2 = 64.692, df = 3, p < 0.001$).

For Japanese listeners, there is a positive relationship between FL and perceptual accuracy in both models. (FL.seg: $\beta = 0.752, p < 0.001$; and FL.sub: $\beta = 0.917, p < 0.001$.) Consistent with the results from the comprehensive models, FL.sub has a greater effect on perceptual accuracy than FL.seg.

The results for English and Korean listeners are conflicting with those from the comprehensive models. In the FL.seg-based model, predicted accuracy is decreasing as FL increases (English:β = -1.518, p < 0.001; Korean; β = -0.826, p < 0.001). In the FL.sub-based model, perceptual accuracy is still decreasing in FL for English listeners (β = -0.470, p < 0.001), but is increasing for Korean listeners (β = 0.888, p < 0.001). The language-specific results are described by Figure 3.7.

3.5 Discussion

The relationship between perceptual accuracy and two distinct information measures, FL and PI, has been studied. Across the three languages and two different definitions of the context (the adjacent consonant (PI.C) and the adjacent vowel (PI.V)), there is generally a negative relationship between PI and perceptual accuracy. This result implies that a

Table 3.7: Estimated coefficients from the FL.seg- (left) and FL.sub-model (right) for onset.

	FL.seg (Onset)	FL.sub (Onset)
ListenerLang, Eng.L	1.037^{***}	0.979^{***}
	(0.361)	(0.296)
ListenerLang, Jpn.L	-1.549^{***}	-0.652^{***}
	(0.331)	(0.252)
Sound_label, t	0.384^{***}	0.343^{***}
	(0.082)	(0.059)
Sound_label, k	0.445^{***}	0.591^{***}
	(0.106)	(0.074)
Vowel, i	0.495^{***}	0.747^{***}
	(0.057)	(0.066)
Vowel, u	0.637^{***}	0.867^{***}
	(0.058)	(0.066)
Kor.L * Kor.T	0.088	0.080
	(0.497)	(0.501)
Eng.L * Kor.T	-0.424	-0.428
	(0.433)	(0.438)
Jpn.L * Kor.T	0.128	0.123
	(0.411)	(0.416)
Kor.L * FL.seg	-0.826^{**}	
	(0.345)	
Eng.L * FL.seg	-1.518^{***}	
	(0.202)	
Jpn.L * FL.seg	0.752^{***}	
	(0.148)	
Kor.L * FL.sub		0.888^{***}
		(0.180)
Eng.L * FL.sub		-0.470^{***}
		(0.160)
Jpn.L * FL.sub		0.917^{***}
		(0.140)
Constant	2.378^{***}	1.418^{***}
	(0.432)	(0.389)
Observations	21,168	21,168
Log Likelihood	$-6{,}190.967$	$-6{,}208.578$
AIC	12,421.930	12,457.160
BIC	12,581.140	12,616.360

Note: Standard errors of the coefficients are given in parentheses.
Coefficients that are significantly different from zero are marked with asterisks.
($^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$)

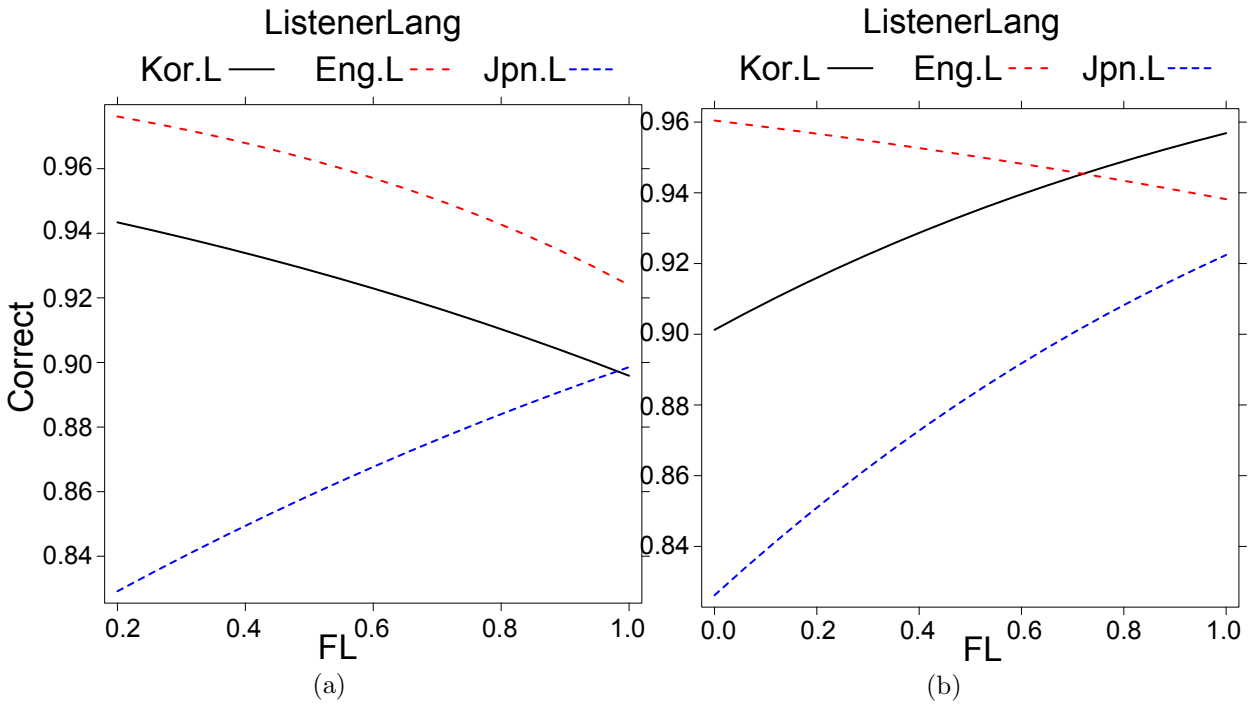

Figure 3.7: Model predictions for the interaction of listener language and (a) FL.seg and (b) FL.sub for onset.

phoneme that is more likely to occur under relevant contexts is more accurately perceived. As defined earlier, contextual familiarity positively correlates with perceptual accuracy.

In contrast, the relationship between perceptual accuracy and FL is positive both for English and Japanese listeners, and only slightly negative for Korean listeners; the coefficient on FL.sub is not significantly different from zero for Korean listeners. This generally positive relationship between FL and perceptual accuracy cannot be interpreted as implying a positive relationship between the amount of information that a phoneme carries and its perceptual accuracy, given the opposite result from PI.

This contrast between PI-based models and FL-based models show that different information measures, even though they all supposedly represent the amount of information in a speech element, can represent very different types of information. This observation is also consistent with the finding that PI and FL are generally negatively related within each of the three languages studied.

Among the three languages, Japanese listeners' perceptual accuracy always shows a consistent relationship with each of the two information measures, FL and PI. With PI, Japanese listeners' perceptual accuracy is more strongly correlated with PI.V than with PI.C, which may be explained by the fact that CV syllables are prevalent in Japanese and hence, a conso-

nant more often neighbors a vowel rather than another consonant. In the onset-specific model with FL, Japanese listeners' perceptual accuracy shows a greater variance with FL.sub (FL computed for CV units) that with FL.seg (FL computed for individual consonants), which may also explained by the fact that CV syllables are prevalent in Japanese; the Japanese listeners may pay a greater attention to CV units that carry a larger amount of information under FL.

Given that such a relationship between experimental results and a syllable structure is only observed for Japanese, the results are far from establishing that such a relationship is universal. However, they at least point to the possibility that, with a further study employing more languages and a larger set of consonants to experiment with, a robust cross-linguistic relationship may be found between a syllable structure and the employment of different information measures in perception.

.

Chapter 4

Experiment 2: Perceptual Accuracy of Audio-Visual Stimuli

4.1 Chapter introduction

The stimuli in Experiment 1 were acoustic waveforms. This chapter's experiment uses audio-visual stimuli to determine how information measures interact with the perception of these richer audio-visual stimuli. The same methods as those used in the last chapter are employed to analyze the perceptual accuracy of different sounds and to compare how the correlations between perceptual accuracy and information measures differ under the two experiment conditions.

The role of visual speech has been acknowledged in the literature. For example, Sumby and Pollack (1954) found that a visual signal can compensate for about 6 dB of noise in speech intelligibility. The McGurk effect (McGurk & MacDonald 1976) indicates that auditory and visual streams of information are integrated with each other at an early stage of speech processing. The importance of visual information is particularly emphasized in language acquisition literature for its exclusive role in face-to-face interaction during acquisition. Without visual input, blind children develop place of articulation contrasts much more slowly than do sighted children (Mills 1983). Evidence that neonates imitate facial gestures also highlights the importance of visual phonetic cues in language acquisition (Meltzoff & Moore 1977).

The patterns of visual similarity among speech sounds, termed "visemes" (Fischer 1968), differ by each speech sound. The similarity is also driven by a number of external factors, such as speaker, listener, and vowel environment, as well as the properties of sounds themselves. Among many, place of articulation is probably one of the most recognizable visual property of all. Jiang et al. (2007) carried out a representative study that showed the result of cluster analyses of sounds that are taken to be visually similar to one another (see Table 4.1). The study showed that, although a stop consonant's place of articulation can generally be identified solely with acoustic input (cf. Miller & Nicely 1953), place contrasts (coronals,

Talker	Vowel	Phoneme Equivalence Classes
M1	ɑ	{w} {m p b} {r f v} {h} {θ ð y l n k g} {t d s z ʃ ʒ ʧ ʤ}
	i	{w} {m p b} {r f v} {y k g h} {}θ ð l n t d } {s z ʃ ʒ ʧ ʤ}
	u	{w y k g h} {m p b} {r f v} {θ ð l n t d} {s z ʃ ʒ ʧ ʤ}
F1	ɑ	{w r} {m p b} {f v} {h} {θ ð} {y l n k g h t d s z ʃ ʒ ʧ ʤ}
	i	{w r} {m p b} {f v} {h} {θ ð} {y l n k g h t d s z ʃ ʒ ʧ ʤ}
	u	{w r m p b} {f v} {h} {θ ð} {y l n k g h t d s z ʃ ʒ ʧ ʤ}
M2	ɑ	{w r} {m p b} {f v} {h} {θ ð} {y l n k g h} {t d s z ʃ ʒ ʧ ʤ}
	i	w r{m p b} {f v}{h}{θ ð}{y l n t k d g h} {s z ʃ ʒ ʧ ʤ}
	u	{w r}{m p b}{f v}{h}{θ ð}{y l n k g h}{t d s z ʃ ʒ ʧ ʤ}
F2	ɑ	{w}{m p b}{r f v}{θ ð y l n k g h t d s z ʃ ʒ ʧ ʤ}
	i	{w}{m p b}{r f v}{θ ð}{y l n k g h}{t d s z}{ʃ ʒ ʧ ʤ}
	u	{m p b}{r f v}{θ ð}{w y l n k g h t d s z ʃ ʒ ʧ ʤ}

Table 4.1: Phoneme equivalence classes obtained by cluster analyses for confusion matrices of each talker and vowel context (adapted from Jiang et al. 2007, p.1075).

labials and dorsals) are also visually very distinct from one another.

The perceptual salience, or sensitivity of stops, is a much studied topic in perception literature in connection with the asymmetric pattern in place assimilation. Specifically, the patterns of perceptual salience have been suggested as the grounds for this asymmetry (Jun 1995, 1996). However, perceptual evidence regarding this asymmetry has been inconsistent, with some authors finding coronal salience to be high, while others found it low (Mohr & Wang 1968, Pols 1983, Winters 2001, 2003, Hura et al. 1992, Wang & Fillmore 1961, Singh & Black 1966, Kochetov & So 2007, Kochetov & Pouplier 2008). Such inconsistent results are partly attributed to various experimental conditions, including the number of talkers in the stimuli (one vs. multiple talkers), the structure of stimuli (isolated syllable vs. clusters), the treatment of the release burst (absence vs. presence of the burst), and listening conditions (see Winters 2001, Kawahara & Garvey 2014 for review).

In addition, the majority of studies focused on the salience in an audio-only condition, with only a few including audio-visual modality in accounting for perceptual salience of stops. Winters (2001, 2003) is one such example that investigated the perceptual salience of stops in audiovisual modality. In an attempt to empirically find a perceptual and articulatory ground for place assimilation, Winters carried out perception experiments with native English listeners in many different conditions using the identification paradigm. The studies found a difference in the perceptual salience of place contrasts of stops and nasals and also a visual advantage most exclusively for labials when audiovisual input is present. Besides the

difference in modality, Winters (2001, 2003) also compared the salience of place contrasts between stops and nasals in various listening conditions and revealed that stops are more salient than nasals only under specific listening conditions. Finally, different talkers used in the experiments caused different salience patterns.

Given the findings from previous studies on the effect of visual stimuli on perception, this chapter describes the results from the perception experiment using audio-visual stimuli and compare them with those from the audio-only experiment described in Chapter 3.

4.2 Method

4.2.1 Video stimuli

The audio part of the experimental stimuli used for the audio-visual condition was identical to those used in the audio-only condition in Chapter 3. Indeed, the recordings used in the audio-only experiment were the soundtracks from the video clips used in this audio-visual experiment.

4.2.2 Video editing

As with the making of the auditory stimuli, a video sequence VC.CV was constructed by concatenating the video recordings of two syllables, VC and CV. For the production of a VC section, when the consonant was /p/, the part where the talkers began moving their lips from the rest position to the point where they clearly made lip closure before making a release burst was carefully inspected and cut for each talker in as consistent way as possible across the talkers. When the consonant was /t/ or /k/, a section of the clip starting from the rest position and ending at a stable lip position right before the release of the burst was used. For the production of a CV section, a section of the clip from the stop onset closure to the vowel articulation was used. The average duration for the edited CV and VC sections was 460 ms. The video recordings and audio recordings aligned with each other very well, and the alignment of each stimulus was checked manually.

The edited CV and CV sections with matching vowels were concatenated using FFMPEG to create 27 audio-visual tokens (3 vowels * 3 onsets * 3 codas) by each talker. Due to the concatenation, there was sometimes a noticeable video transition between VC and CV sections in some of the edited videos, which was caused by a slight change in head position or an unnatural lip movement, for example. When the transition was too abrupt, the original clips of the VC and CV syllables were edited further or a different recording of the same syllable was selected to create a more natural-looking clip. In the end, all of the clips looked relatively natural.

4.2.3 Participants

A total of 43 English, Japanese and Korean listeners participated in the experiment. Participants who had not taken part in the audio-only experiment were recruited. English listeners were recruited from University of California, Berkeley; Korean listeners from Seoul National University in Seoul, South Korea; and finally, Japanese listeners from Daito Bunko University in Tokyo, Japan. The recruitment guidelines and conditions followed those used in the audio-only experiment. Table 4.2 gives a summary of the participants.

Language	Male	Female	Total
English	6	7	13
Korean	8	7	15
Japanese	5	10	15

Table 4.2: Summary of the participants from the three language groups.

The participants were native speakers of English, Japanese and Korean, and reported no hearing problems. Some native listeners of English were relatively fluent in other languages such as Spanish or French, but none of them had been exposed to Japanese or Korean. Also, no Korean or Japanese listeners had spent more than a year in an English-speaking country before the age of 13.

4.2.4 Procedure

The identification paradigm used in the audio-only experiment was replicated in the audio-visual experiment. The experiment took place in the university from which each listener was recruited. Although the sites of the experiment were different for the listeners depending on their native languages, the same procedure was employed for all the three language groups in running the experiment. The participants were first given both verbal and written instructions, and later seated in front of a computer screen in a sound-attenuated room. After the session started, the experiment stimuli were presented by Opensesame software (Mathôt et al. 2012). Each video clip was played one at a time, at the end of which the listeners were asked to identify the intervocalic clusters they had just watched. Throughout the session, the participants wore AKG K 240 headphones and were given a volume switch, with which they could freely adjust the volume to their comfort level.

The participants were explicitly asked to watch the video during the entire session. Since the task could technically be completed by only listening to the soundtracks of the videos without watching the screen, it was possible for the listeners not to pay attention to the screen; such a behavior would make the experiment condition identical to that of the audio-only experiment. In order to avoid this, the listeners were asked to pay close attention to the computer screen as they completed the task.

To make sure the listeners watch the screen for the whole session, 24 'catch' trials were randomly inserted among the 216 audio-visual tokens. Each catch trial consisted of a video clip taken from one of the 216 tokens and a yellow subtitle saying "do not respond," positioned at approximately 1/4 of the way from the bottom of the screen. The participants were instructed to click on a 'pass' button when they saw such a subtitle. Otherwise, the participants identified the CC in the video clip as one of the nine alternatives, which represented all the possible two-consonant sequences of /p/, /t/ and /k/. In addition, they had the option of clicking on 'not sure.' In total, the participants were given 11 choices, and the next video clip was played once they have clicked on one of the choices with a mouse.

4.2.5 Analysis

The same statistical methods as those used in Chapter 3 were employed to construct models for both PI and FL. Both of PI and FL were computed using the same corpora and techniques as in Chapter 3.

Only the listeners who correctly clicked on 'pass' after seeing a catch trial at least 50 percent of the time were included in the final analysis. This treatment is necessary to make sure that only those participants who paid close attention to the visual token are taken into account.

4.3 Results and discussion

4.3.1 Correct response rate to catch trials

The 'correct' answer to a catch trial is defined as 'pass.' The average correct response rate to a catch trial across all of the 43 participants is 87.11%. It is the highest for the Korean listeners (96.36%), followed by the English listeners (87.19%). The Japanese listeners show the lowest correct response rate, which is 76.94%. Specifically, there are two Japanese listeners with 0% correct response rate, and two other Japanese listeners with relatively low rates of 58% and 66%. All the listeners other than these four show correct response rates higher than 75%.

Therefore, only the two Japanese listeners with 0% correct response rate to a catch trial are dropped from the analysis. The remaining 13 English, 13 Japanese and 15 Korean listeners' responses constitute the data points for the statistical analysis.

4.3.2 Informativity

4.3.2.1 Model estimation

The models have the same controlling factors and PI measures as those in the previous chapter. As a reminder, two PI measures used: PI.V, which takes the adjacent vowel as a

context, and PI.C, which takes the adjacent consonant as a context. The models can be expressed as follows:

$$Prob(Correct = 1) = P[\alpha_1 + \alpha_2 + \sum_x \beta_x PI \cdot I(ListenerLang = x) + \mathbf{\Gamma} \cdot (\text{Control Variables})]. \tag{4.1}$$

This equation represents a mixed effects logistic regression. P is the cumulative distribution function of a logistic distribution. α_1 and α_2 are subject and talker random intercepts, respectively, and $\mathbf{\Gamma}$ is the vector of coefficients on the control variables, which are ListenerLang, Syll.pos, Sound_label, Vowel, NearC, and talker language–ListenerLang interaction. No random slope is included; with random slopes, neither the PI.V- nor the PI.C-based model converges.

The coefficient β represents the effects of PI by each listener language group, which is the coefficient of main interest. The variable x represents the three languages, English, Japanese and Korean. Table 4.3 shows the estimated coefficients of the models.

Control factors The effects of the controlling factors are graphically illustrated in Figure 4.1 and 4.2 for PI.V and PI.C, respectively.

The overall pattern is similar between the two models. Among the controlling factors, the significant effect of the consonant type, Sound_label, is worth noting. Unlike the audio-only experiment reported in Chapter 3, /p/ is associated with the highest predicted accuracy in this audio-visual experiment. This result is consistent with the availability of visual information since labials are visually more prominent than other stops (e.g. Winters 2001). A log-likelihood test shows that the effects of Sound_label are significant (PI.V-model: $\chi^2 = 500.98, df = 2, p < 0.001$; and PI.C-model:$\chi^2 = 302.05, df = 2, p < 0.001$).

In addition, adjacent vowels also significantly affect perception: In both models, listeners' accuracy is lower for stops adjacent to /i/ than those adjacent to /a/ and /u/. (PI.V: $\chi^2 = 8.154, df = 2, p = 0.017$; PI.C: $\chi^2 = 8.166, df = 2, p = 0.017$.) Adjacent consonants also significantly affect the responses. (PI.V: $\chi^2 = 68.808, df = 5, p < 0.001$; PI.C: $\chi^2 = 55.385, df = 5, p < 0.001$.)

Critical factors The estimated models indicate a significant language-specific effect of PI on predicted accuracy, as shown by Figures 4.3 and 4.4.

For Japanese and Korean listeners, perceptual accuracy is negatively correlated with both PI.V and PI.C; this result is consistent with that from the audio-only experiment. However, for Japanese listeners, the effect is significant only in the PI.C-model. (PI.V-model: β=-0.279, p = 0.100; PI.C-model: β = -0.462, p < 0.001.) For Korean listeners, the effect is significant only in the PI.V-model. (PI.V-model: β=-0.242, p < 0.001; PI.C-model: β = -0.166, p = 0.356.)

For English listeners, the correlation is positive both in the PI.V-model and the PI.C-model, which contrasts with the results for the other two language groups. As a reminder,

Table 4.3: Estimated coefficients from the PI.C- (left) and PI.V-model (right).

	PI.V-model	**PI.C-model**
ListenerLang, Eng.L	-1.009^{***}	-0.814^{***}
	(0.229)	(0.250)
ListenerLang, Jpn.L	-0.306^{**}	-0.166
	(0.120)	(0.119)
Syll.pos, Onset	2.321^{***}	2.147^{***}
	(0.074)	(0.066)
Sound_label, t	-0.434^{***}	-0.290^{***}
	(0.065)	(0.066)
Sound_label, k	-1.334^{***}	-1.150^{***}
	(0.063)	(0.074)
Vowel, i	-0.139^{***}	-0.139^{***}
	(0.051)	(0.051)
Vowel, u	-0.028	-0.028
	(0.052)	(0.052)
NearC, p	-0.113^{**}	-0.113^{**}
	(0.052)	(0.052)
NearC, t	-0.083	-0.083
	(0.052)	(0.052)
Kor.L*Kor.T	0.883^{***}	0.870^{***}
	(0.300)	(0.300)
Eng.L*Kor.T	0.309	0.315
	(0.298)	(0.298)
Jpn.L*Kor.T	-0.020	-0.021
	(0.296)	(0.296)
Kor.L*PI	-0.242^{**}	-0.166
	(0.112)	(0.180)
Eng.L*PI	0.911^{***}	0.662^{***}
	(0.130)	(0.145)
Jpn.L*PI	-0.279	-0.462^{***}
	(0.170)	(0.104)
Constant	1.463^{***}	1.339^{***}
	(0.251)	(0.253)
N	17712	17712
Log Likelihood	-7176.918	-7183.630
AIC	14389.840	14403.260
BIC	14529.910	14543.330

Note: Standard errors of the coefficients are given in parentheses.
Coefficients that are significantly different from zero are marked with asterisks.
($^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$)

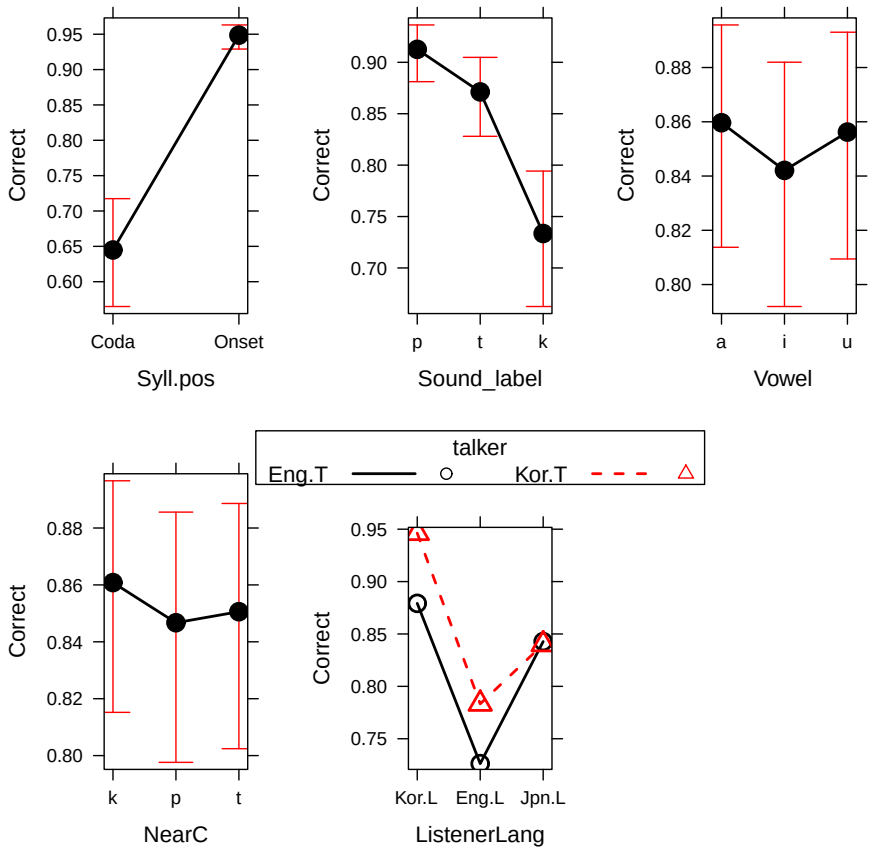

Figure 4.1: Model predictions for selected control factors for the PI.V-model: Syllable position, the adjacent vowel and consonant, and sound.

in the audio-only experiment, English listeners showed an inconsistent correlation across the two models, which was negative in the PI.V-model but positive in the PI.C-model.

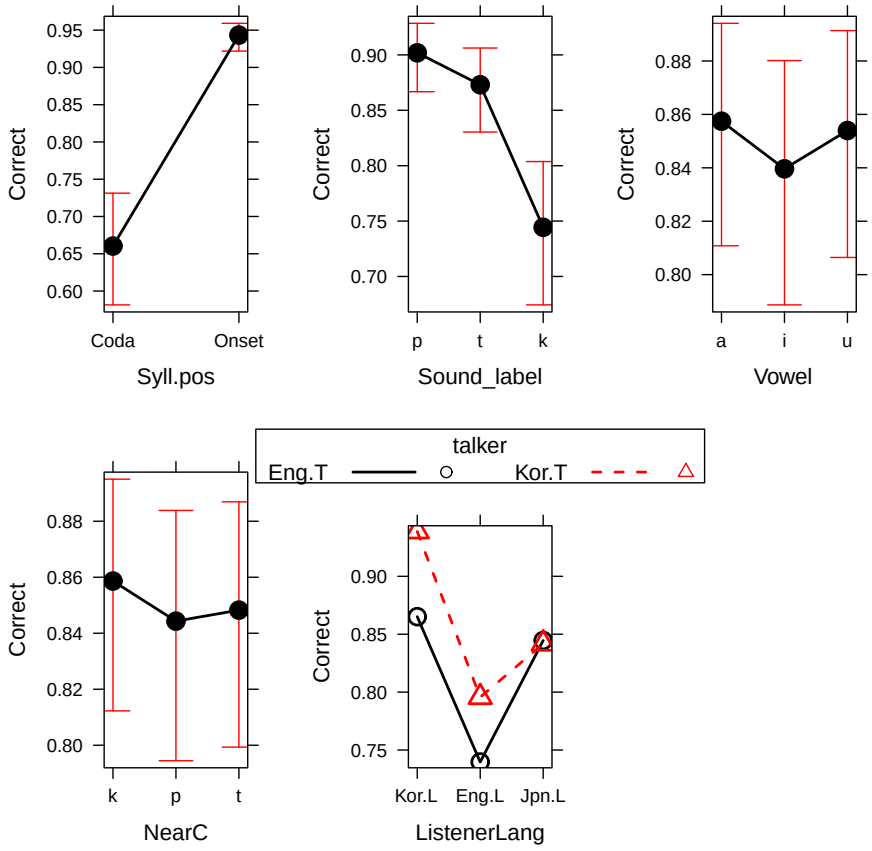

Figure 4.2: Model predictions for selected control factors for the PI.C-model: Syllable position, the adjacent vowel and consonant, and sound.

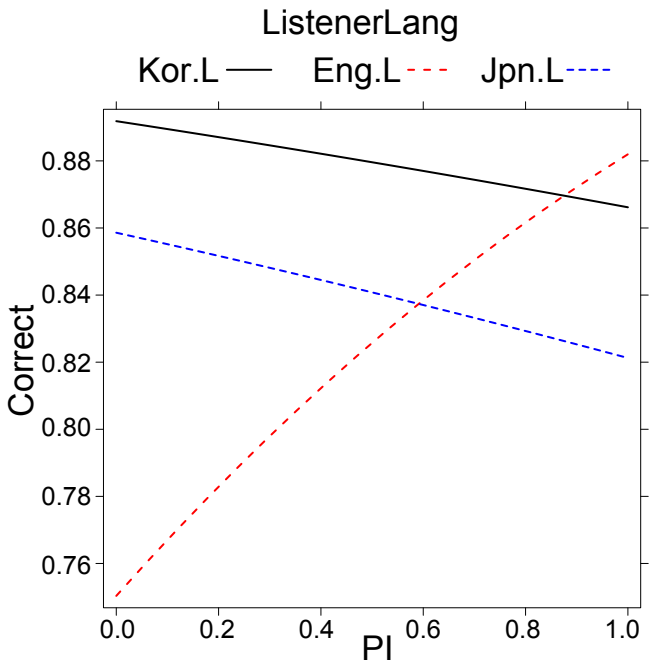

Figure 4.3: Model predictions for PI.V for each listener language group.

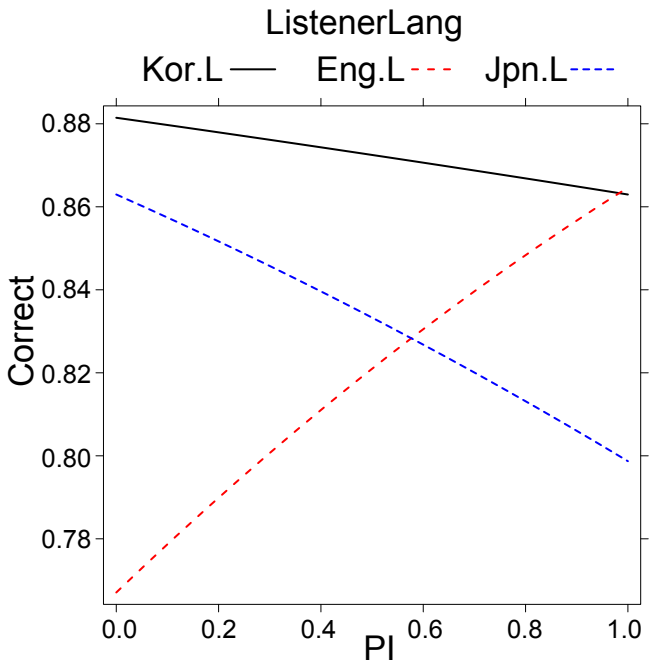

Figure 4.4: Model predictions for PI.C for each listener language group.

4.3.3 Functional load

4.3.3.1 Overview

This section discusses how perceptual accuracy changes as a function of the two types of FL measures: FL.seg, which is computed for individual consonants, and FL.sub, which is computed for VC and CV units. This section employs statistical models identical to those used in analyzing the audio-only experiment.

4.3.3.2 Comprehensive models

Model estimation The same mixed effects logistic regression model is used as that used for PI. With FL.seg, all of the control factors are included. Since random slopes do not improve the model fit, only random intercepts are included.

With FL.sub, NearC is dropped from the set of control factors; otherwise, the regression model does not converge. Random slopes do not improve model fit for the FL.sub-model either, so only random intercepts are included. Table 4.4 lists the estimated coefficients.

Control factors Differently from the audio-only experiment, /p/ is associated with the highest predicted accuracy. This result is consistent with the availability of visual information since labials are visually more prominent than other stops (e.g. Winters 2001). A log-likelihood test shows that the effects of Sound_label are significant (PI.V-model: $\chi^2 = 500.98, df = 2, p < 0.001$; and PI.C-model:$\chi^2 = 302.05, df = 2, p < 0.001$).

The estimated coefficient on Sound_label indicates that /p/ is more accurately perceived than /k/ and /t/. PI models produced the same results. These consistent results across models with different information measures strengthen the argument that the higher accuracy of /p/ may be due to the advantages of the visual lip rounding that occurs with /p/.

In addition, stops adjacent to /i/ show lower perceptual accuracy than those adjacent to /a/ and /u/ in both models, which was also observed with PI models.

Critical factors For Japanese listeners, both FL.seg and FL.sub have a significant and positive effect on perceptual accuracy, which is consistent with the findings from the audio-only experiment. For Korean listeners, only the negative relationship between perceptual accuracy and FL.sub is significant, and no significant relationship between perceptual accuracy and FL is found for English listeners.

Table 4.4: Estimated coefficients from the FL.seg- (left) and FL.sub-model (right).

	FL.seg-model	FL.sub-model
ListenerLang, Eng.L	-0.394^{*}	-0.428^{**}
	(0.216)	(0.213)
ListenerLang, Jpn.L	-0.932^{***}	-0.773^{***}
	(0.107)	(0.102)
Syll.pos, Onset	1.739^{***}	1.891^{***}
	(0.075)	(0.057)
Sound_label, t	-0.547^{***}	-0.469^{***}
	(0.062)	(0.056)
Sound_label, k	-1.339^{***}	-1.297^{***}
	(0.060)	(0.055)
Vowel, i	-0.139^{***}	-0.127^{**}
	(0.051)	(0.054)
Vowel, u	-0.028	-0.038
	(0.052)	(0.058)
NearC, p	-0.114^{**}	
	(0.052)	
NearC, t	-0.083	
	(0.052)	
Kor.L*Kor.T	0.893^{***}	0.883^{***}
	(0.301)	(0.299)
Eng.L*Kor.T	0.316	0.322
	(0.298)	(0.297)
Jpn.L*Kor.T	-0.020	-0.021
	(0.297)	(0.295)
Kor.L*FL.seg	-0.132	
	(0.122)	
Eng.L*FL.seg	0.031	
	(0.175)	
Jpn.L*FL.seg	1.708^{***}	
	(0.189)	
Kor.L*FL.sub		-0.353^{**}
		(0.150)
Eng.L*FL.sub		0.131
		(0.205)
Jpn.L*FL.sub		1.530^{***}
		(0.206)
Constant	1.736^{***}	1.604^{***}
	(0.252)	(0.252)
N	17,712	17,712
Log Likelihood	−7,155.382	−7,176.518
AIC	14,346.760	14,385.030
BIC	14,486.840	14,509.550

Note: Standard errors of the coefficients are given in parentheses.
Coefficients that are significantly different from zero are marked with asterisks.
($^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$)

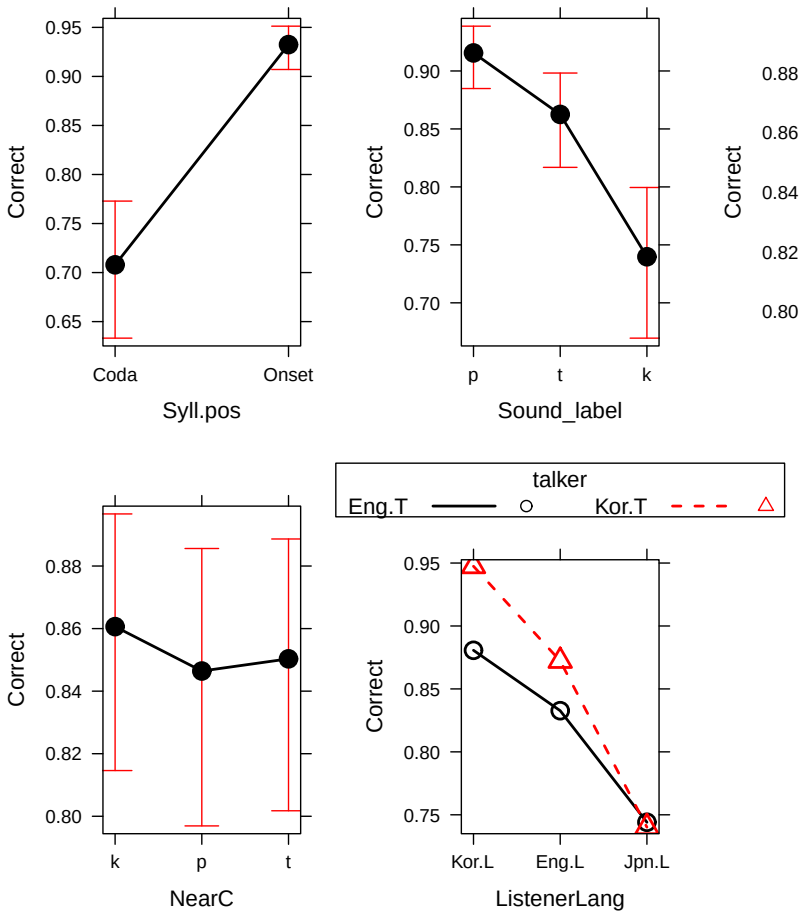

Figure 4.5: Model predictions for selected control factors for the FL.seg-model: Syllable position, the adjacent vowel and consonant, and sound.

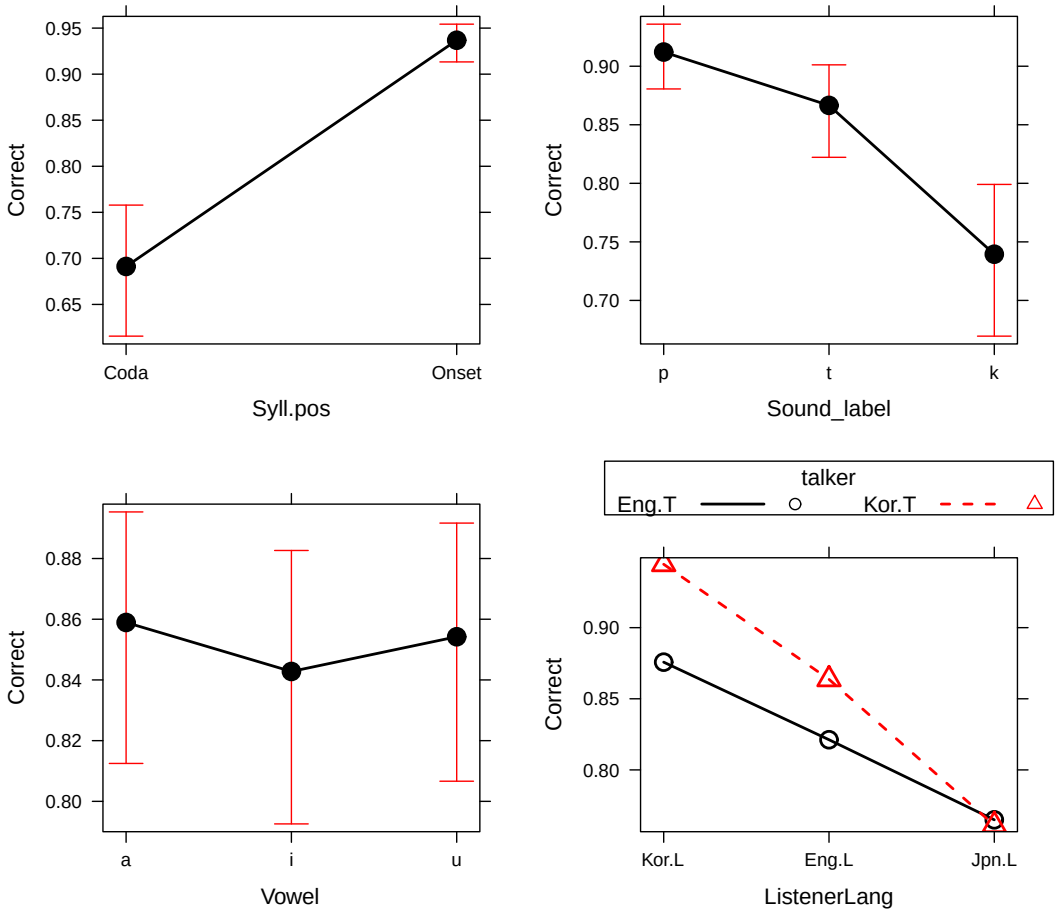

Figure 4.6: Model predictions for selected control factors for the FL.sub-model: Syllable position, the adjacent vowel and consonant, and sound.

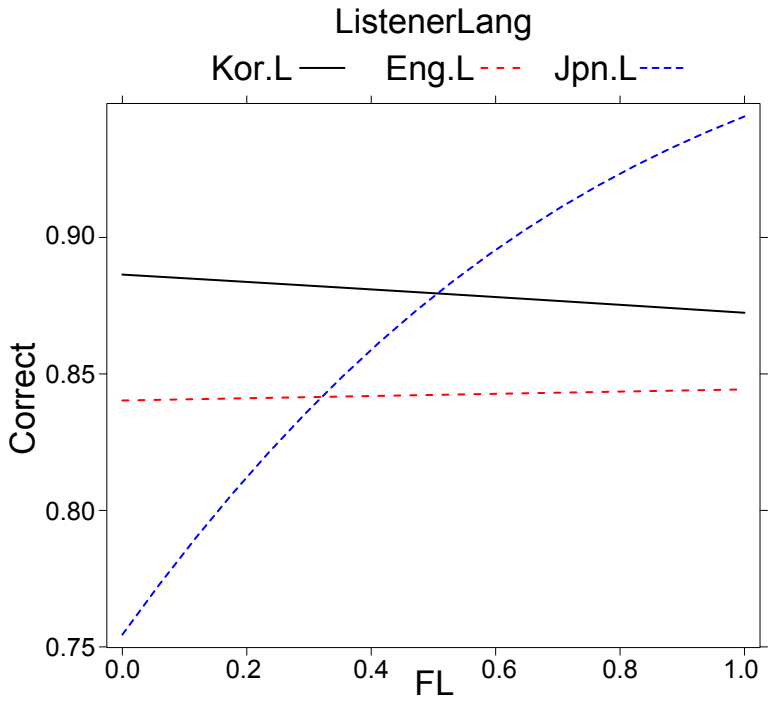

Figure 4.7: Model predictions for FL.seg for each listener language group.

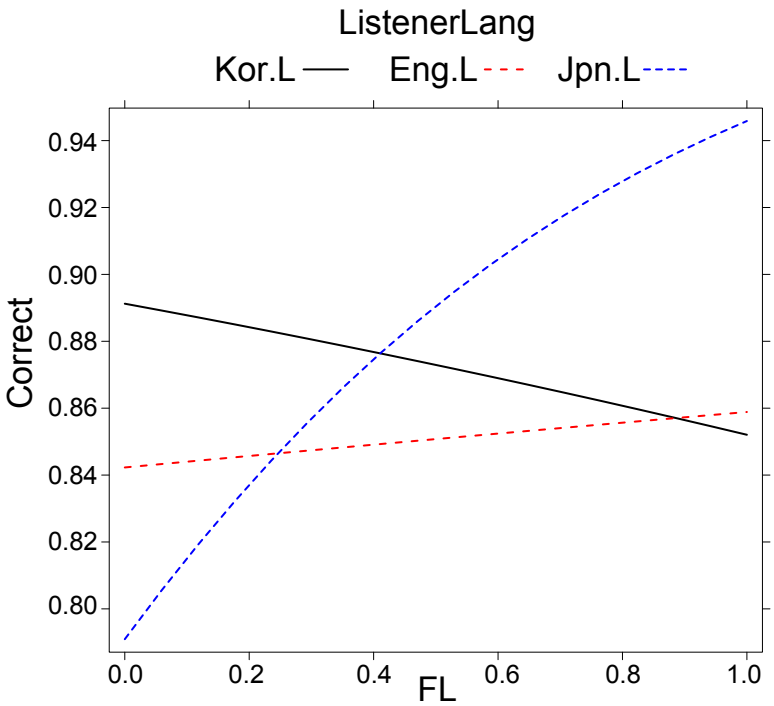

Figure 4.8: Model predictions for FL.sub for each listener language group.

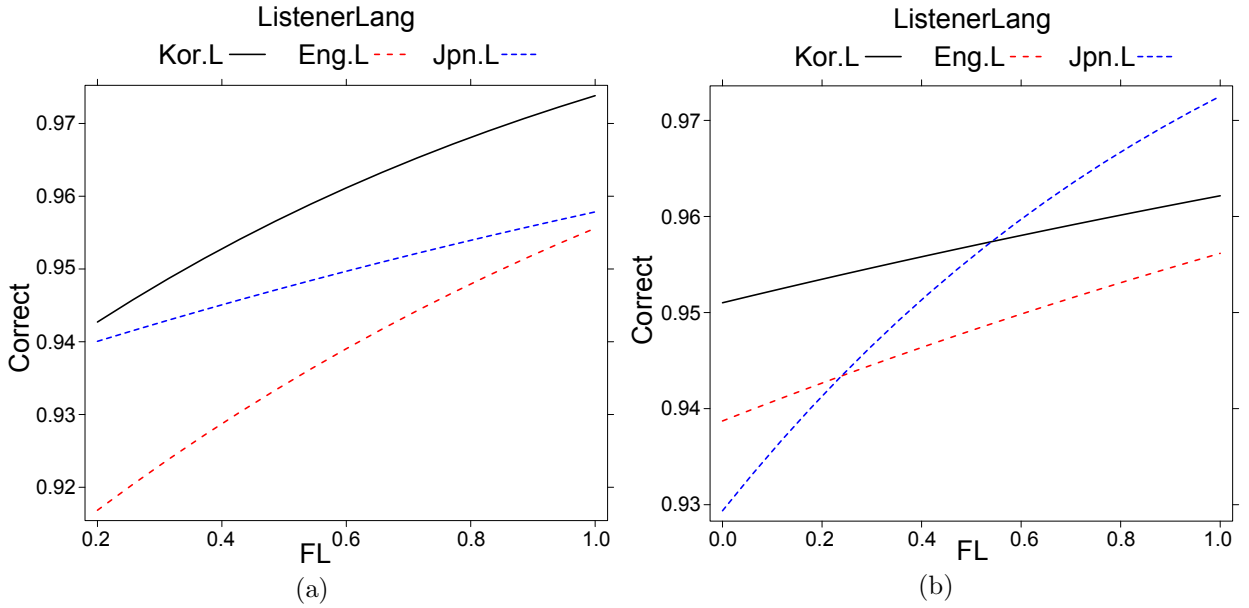

Figure 4.9: Model predictions for (a) FL.seg and (b) FL.sub for onset.

4.3.3.3 Onset-specific models

As with the audio-only experiment, onset-specific models are estimated, which are mixed effects logistic regression models identical to the comprehensive models. SYLL.POS is no longer a part of the model because only responses to onsets are used. The FL.seg-model only includes random intercepts but the FL.sub-model includes random slopes on LISTENERLANG as they improve the model fit. ($\chi^2 = 20.865, df = 10, p = 0,022.$)

Table 4.5 show the estimated coefficients from the two models. The control factors generally elicit similar patterns to those found in the comprehensive models. However, sound /p/ does not have higher perceptual accuracy than /t/, suggesting a possible ceiling effect in onset position given the very high rate of correct responses from the experiment.

For Japanese listeners, the effect of both FL.seg and FL.sub on perceptual accuracy is positive and significant, which is consistent with both what was found with the audio-only experiment and with the comprehensive models for the audio-visual experiment. (FL.seg: β = 0.463, p $<$ 0.001; FL.sub: β = 0.717, p = 0.042.) For English and Korean listeners, the effect is positive, but not always significant. Figure 4.9 plots these effects.

Table 4.5: Estimated coefficients for the FL.seg- (left) and FL.sub-model (right).

	FL.seg-model	FL.sub-model
ListenerLang, Eng.L	−0.364	−0.237
	(0.452)	(0.428)
ListenerLang, Jpn.L	0.063	−0.389
	(0.470)	(0.294)
Sound_label, t	−0.078	0.091
	(0.147)	(0.116)
Sound_label, k	−0.198	−0.395***
	(0.153)	(0.118)
Vowel, i	0.620***	0.746***
	(0.102)	(0.119)
Vowel, u	0.763***	0.886***
	(0.106)	(0.120)
NearC, p	0.049	0.049
	(0.109)	(0.110)
NearC, t	−0.303***	−0.305***
	(0.103)	(0.104)
Kor.L * Kor.T	0.992***	1.003***
	(0.294)	(0.348)
Eng.L * Kor.T	0.389	0.386
	(0.274)	(0.254)
Jpn.L* Kor.T	0.685**	0.717**
	(0.284)	(0.352)
Kor.L * FL.seg	1.019**	
	(0.502)	
Eng.L * FL.seg	0.836**	
	(0.376)	
Jpn.L * FL.seg	0.463*	
	(0.261)	
Kor.L * FL.sub		0.270
		(0.289)
Eng.L * FL.sub		0.353
		(0.288)
Jpn.L * FL.sub		0.989***
		(0.254)
Constant	1.961***	2.250***
	(0.410)	(0.307)
Observations	8,856	8,856
Log Likelihood	−2,033.128	−2,020.084
AIC	4,100.256	4,094.169
BIC	4,220.767	4,285.567

Note: Standard errors of the coefficients are given in parentheses.
Coefficients that are significantly different from zero are marked with asterisks.
($^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$)

4.4 Discussion

The comparison of perceptual accuracy between the three stop sounds provides evidence that listeners are sensitive to visual cues. In both PI-based and FL-based models, the stop /p/, which is more visually prominent than /t/ and /k/ (e.g. Winters 2001), shows the highest predicted perceptual accuracy. With vowels, stops adjacent to /i/ are less accurately perceived than those adjacent to /a/ and /u/, but there is no clear explanation for this result.

For Japanese listeners, the results from the audio-only experiment and the audio-visual experiments are highly consistent: perceptual accuracy is negatively correlated with PI in both experiments, but is positively correlated with FL in both experiments.

Korean listeners also show a negative relationship between perceptual accuracy and PI in both experiments. They exhibit inconsistent results with FL, but this inconsistency may reflect a ceiling effect, given the generally high rate of correct responses by Korean listeners, especially in onset position.

English listeners produce inconsistent results with both PI and FL, within each of the two experiments as well as across them. These results exhibit an apparent language-specificity in that information measures reliably predict perceptual accuracy in some languages, but not in others. However, ceiling effects might be driving some of the inconsistency, especially in onset position; a more difficult task with a correct response rate far from 100% may be able to capture a robust correlation across languages.

Chapter 5

Conclusion

The first section briefly summarize the findings from the previous chapters. Discussions of their significance, theoretical implications, and limitations follow.

5.1 Summary and significance of the findings

This thesis has investigated the relationship between perceptual accuracy and information measures and the cross-linguistic variation of that relationship. Experiments were conducted to measure the perceptual accuracy for the three stop sounds /p/, /t/ and /k/ with three listener groups whose native languages were English, Japanese and Korean.

Chapter 2 described a corpus analysis on the three languages, English, Japanese and Korean. A lexical corpus for each language was transformed to construct a syllable corpus, from which two information measures, PI and FL, were computed. For both English and Korean, the mean of PI for consonants in onset position was not significantly different from the mean of PI in coda position. In Japanese, the mean of PI for consonants was significantly higher in coda than in onset.

The comparison of FL across different syllable positions and languages yielded a different pattern. Even though both PI and FL represent the amount of information that a speech element carries, the quantity of information can be measured very differently depending on how it is defined; the differences observed between PI and FL confirmed this conventional wisdom.

For all the three languages, the mean of FL for consonants was higher in onset position than in coda position, and the mean difference was significant for English and Japanese. The mean difference was the largest for Japanese, which may be driven by the fact that CV syllables are prevalent in Japanese, thus codas do not differentiate as many syllable occurrences as onsets do. The comparison of mean FL between CV and VC units yielded the same result; mean FL is higher for VC units than for CV units in all the three languages and the difference is largest for Japanese.

Also, the mean FL of consonants in coda position for Korean was significantly higher than those for English and Japanese. This result is consistent with the fact that (a) in Japanese and Korean, there are much fewer types of codas than in English; and (b) in Korean, syllables with a coda are much more common than in Japanese.

Chapter 3 analyzed the responses gathered in the audio-only experiment to measure the relationship between perceptual accuracy and the two types of information measures, PI and FL. A generally negative relationship between perceptual accuracy and PI was found across the three languages and the two different definitions of contexts in computing PI; PI.V was computed with the adjacent vowel as the context, and PI.C was computed with the adjacent consonant as the context. This result implies that a phoneme that is more likely to occur in a particular context (contextual familiarity) is more accurately perceived.

In contrast, the relationship between perceptual accuracy and FL was positive both for English and Japanese listeners, and only slightly negative for Korean listeners.

Between the three languages, Japanese listeners' perceptual accuracy always showed a consistent relationship under each of the two information measures, PI and FL. Their perceptual accuracy was more strongly correlated with PI.V than with PI.C, which may be explained by the fact that CV syllables are prevalent in Japanese and hence, a consonant more often neighbors a vowel rather than another consonant. In the onset-specific model, Japanese listeners' perceptual accuracy showed a greater variation with FL.sub (FL computed for CV units) than with FL.seg (FL computed for individual consonants), which may also explained by the fact that CV syllables are prevalent in Japanese; Japanese listeners may pay a greater attention to CV units that carry a larger amount of information under FL.

Chapter 4 analyzed the responses gathered in the audio-visual experiment. The additional visual stimuli apparently had a significant impact on the listeners; the stop /p/, which is more visually prominent than /t/ and /k/ (e. g. Winters 2001), showed the highest predicted perceptual accuracy, under both PI-based and FL-based models.

For Japanese listeners, the results from the audio-only experiment and the audio-visual experiments were very close to each other; the two experiments implied identical signs of correlation between perceptual accuracy and the information measure under every model specification. For English and Korean listeners, the results were less consistent between the two experiments.

The observed generally negative correlation between PI and perceptual accuracy may be counterintuitive, but can be explained in two ways. First, listeners may be simply perceiving contextually familiar sounds more accurately. The other explanation is from the enhancement account of a listener-oriented sound change. According to Garrett and Johnson (2011), the enhancement account argues that listeners tend to perceive an ambiguous auditory signal as a phone that is likely to be mis-produced. There is no known relationship between PI and the probability of mis-production, but Cohen Priva (2008) found a negative relationship between PI and the rate of sound deletion. Likewise, if PI and the probability of mis-production were negatively correlated, listeners would tend to perceive an ambiguous sound as a phone with low PI. In the experiments, this behavior would increase the rate of

a correct response to a phone with low PI and decrease the correct response rate to a phone with high PI.

5.2 Limitations of the current study

There are shortcomings in the experiment design and its implementation. First, while Japanese listeners also participated in the perception experiment, only English and Korean talkers recorded the stimuli. Although the effects of any same-language talker advantage (or disadvantage) were carefully controlled for by including talker-listener language interaction as a fixed factor, the Japanese listeners still may have suffered the double disadvantages from foreign talkers and foreign sequences.

Another caveat is that only three sounds were used in the experiment. This may be responsible for not only the failure of some models to converge, but also for some unusual variations and extreme ranges in predicted accuracy in some models. This limitation is inevitable to some extent, as the experiment stimuli were confined to sounds that commonly occur in all the three languages. However, since these languages share more than three consonants, a further study that includes a greater variety of sounds may reinforce the results of this thesis.

5.3 Final remark

This thesis has explored the relationship between information measures of different speech elements and perceptual accuracy. The current thesis is the first to provide empirical evidence that different definitions of information measures can produce quite different numbers, which can be even negatively correlated. In addition, information measures are shown to have effects on perceptual accuracy, which vary significantly across different languages and different information measures. The statistical properties of a corpus are useful in predicting perceptual behavior, but the cross-linguistic variations show that such a prediction must be based on characteristics of a language, such as its syllable structure. For example, the results from Japanese listeners suggest that the syllable structure of a listener's native language may determine the granularity of speech that he or she is attuned to. Altogether, this thesis presents additional evidence on the close relationship between a listeners' language background and speech perception.

Reference

Ahn, S.-C., & Iverson, G.K. (2007). Structured Imbalances in the Emergence of the Korean Vowel System. In *Historical Linguistics* (pp. 275-294), ed. Joseph C. Salmons and Shannon Dubenion-Smith. Amsterdam: John Benjamins.

Amano, S., & Kondo, T. (1999, 2000). *Lexical Properties of Japanese* Tokyo:Sanseido (in Japanese)

Baayen R. H. (2008).*Analyzing linguistic data: A practical Introduction to Statistics Using R.* Cambridge: Cambridge University Press.

Barr, D. J., Levy, R., Scheepers, C., & Tilly, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*, 255-278.

Bates, D., Maechler, M., Bolker, B. & Walker, S. (2015). lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1-9, https://CRAN.R-project.org/package=lme4.

Best, C., Strange, W. (1992). Effects of phonological and phonetic factors on cross-language perception of approximants, *Journal of phonetics 20*, 305330.

Bergier, M. (2014). Influence of explicit phonetic instruction and production training practice on awareness raising in the realization of stop consonant clusters by advanced polish learners of English. In Lyda, A. & Szcześniak, K.(Eds.), *Awareness in Action; The Role of Consciousness in Language Acquisition* (pp. 103-120), Springer International Publishing Switzerland.

Best, C. T., McRoberts, G.W., Lafleur, R., & Silverisenstadt, J. (1995). Divergent developmental patterns for infants' perception of two nonnative speech contrasts. *Infant Behavior and Development, 18*, 339-350.

Boersma, P., & Weenink, D. (2007). Praat: doing phonetics by computer [Computer program]. Version 4.5.25, http://www.praat.org/.

Broselow, E., & Finer, D. (1991). Parameter setting in second language phonology and syntax, *Second Language Research, 7*, 35-59.

Brysbaert, M., & New, B. (2009). Moving beyond Kucera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behav Res Methods 41*: 977990.

Chang, C. B. (2014). Bilingual perceptual benefits of experience with a heritage language. *Bilingualism: Language and Cognition*, available on CJO2014. doi:10.1017/S1366728914000261.

Cho, T., Jun, S.-A., & Ladefoged, P. (2002). Acoustic and aerodynamic correlates of Korean stops and fricatives. *Journal of Phonetics, 30*(2), 193-228.

Choi, H. (2002). Acoustic cues for the Korean stop contrast – dialectal variation. *ZAS Papers in Linguistics, 28*, 1-12.

Cohen Priva, U. (2008). Using information content to predict phone deletion. In N. Abner, J. Bishop, (Eds.), *Proceedings of the 27th West Coast Conference on Formal Linguistics* (pp. 90-98).

Cohen Priva, U. (2012). Sign and signal: deriving linguistic generalizations from information utility. PhD dissertation, Stanford University.

Cutler, A, Mehler, J., Norris, D., & Segui, J. (1986). The syllable's differing role in the segmentation of French and English. *Journal of Memory and Language, 25*(4), 385-400.

Cutler, A., Mehler, J., Norris, D., & Segui, J. (1987). Phoneme identification and the lexicon. *Cognitive Psychology, 19*(2), 141-177.

Dahaene-Lambertz, G., Dupoux, E., & and Gout, A. (2000). Electrophysiological correlates of phonological processing: A cross-linguistic study. *Journal of Cognitive Neuroscience, 12*, 635-647.

Davidson. L., (2006). Phonology, phonetics, or frequency: Influence on the production of non-native sequences, *Journal of Phonetics, 34*, 104-137.

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

Diehl, R. L., Kluender, K. R., Foss, D. J., Parker, E. M., & Gernsbacher, M. A. (1987). Vowels as islands of reliability. *Journal of Memory and Language, 26*, 564-573.

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*, 1568-78.

Eckman, F., & Iverson, G. (1993). Sonority and markedness among onset clusters in the interlanguage of ESL learners. *Second Language Research, 9*, 234-252.

Elman, J. L., & McClelland, J. L. (1988). Cognitive penetration of the mechanisms of perception: Compensation for coarticulation of lexically restored phonemes. *Journal of Memory and Language, 27*(2), 143-165.

Fisher, C., (1968). Confusions among visually perceived consonants, *J. Speech Hearing Research, 11*, 796-804.

Fisher, W., (1996). Tsylb syllabification package. ftp://jaguar.ncsl.nist.gov/pub/tsylb2-1.1.tar.Z. Last accessed 10 December 2015.

Flege, J. E., & Wang. (1989). Native-language phonotactic constraints affect how well Chinese subjects perceive the word-final /t/-/d/ contrasts. *Journal of Phonetics, 17*, 299-315.

Foss, D., & Swinney, D. (1973). On the psychological reality of the phoneme: Perception, identification, and consciousness. *Journal of Verbal Learning and Verbal Behavior, 12*(3), 246-257.

Galantucci, B., Fowler, C. A., & Turvey, M. T. (2006). The motor theory of speech perception reviewed. *Psychonomic Bulletin & Review, 13*(3), 361-377.

Garrett, A., & Johnson, K. (2013). Phonetic bias in sound change, A.C.L. Yu (Ed.), *Origins of Sound Change: Approaches to Phonologization* (pp. 51-97), Oxford: Oxford University Press.

Gimson, A. C., & Cruttenden, A. (1994). *Gimson's Pronunciation of English.* London: Oxford University Press Inc.

Granath, S. (2007). Size matters - Or thus can meaningful structures be revealed in large corpora. In R. Facchinetti (ed.), *Corpus Linguistics 25 Years On* (pp. 169-185). Amsterdam, the Netherlands:Rodopi.

Gamkrelidze, T. V. (1978). On the correlation of stops and fricatives in phonological system. In J. Greenberg (ed.), *Universals of Human Language vol. 2, Phonology*(pp.9-46) Stanford University Press.

Goldrick, M. (2002). Patterns of sound, patterns in mind: Phonological regularities in speech production. PhD dissertation, Johns Hopkins University.

Hall, N. (2011). Vowel epenthesis. In Marc van Oostendorp, C. J. Ewen, E. Hume & K. Rice (eds.) *The Blackwell companion to phonology.* Malden, MA & Oxford: Wiley-Blackwell. pp.1576-1596.

Höhle, B., Bijeljac-Babic, R., Herold, B., Weissenborn, J., & Nazzi, T. (2009). Language specific prosodic preferences during the first half year of life: evidence from German and French infants. *Infant Behavior & Development, 32*(3), 262-74.

Henderson, J.B., & Repp, B.H. (1982). Is a stop consonant released when followed by another stop consonant? *Phonetica, 39*, 71-82.

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience, 8*(5), 393-402.

Hlavac, M. (2015). stargazer: Well-Formatted Regression and Summary Statistics Tables. R package version 5.2. http://CRAN.R-project.org/package=stargazer

Hockett, C. F. (1967). The quantification of functional load: A linguistic problem. Word, 23: 320-339.

Hon, N., & Tan, C. (2013). Why rare targets are slow: Evidence that the target probability effect has an attentional locus. *Attention, Perception, & Psychophysics.* Retrieved from http://link.springer.com/article/10.3758/s13414-013-0434-0

Houston, D. M., Jusczyk, P.W., Kuijpers, C., Coolen, R., & Cutler, A. (2000). Cross-language word segmentation by 9-month-olds. *Psychonomic Bulletin and Review, 7*(3), 504-509.

Huang, T., & Johnson, K. (2010). Language specificity in speech perception: Perception of Mandarin tones by native and non-native speakers. *Phonetica, 67*, 243-267.

Hura, S., Lindblom, B., & Diehl, R. (1992). On the role of perception in shaping phonological assimilation rules. *Language and Speech, 35*, 59-72.

Jacquemot, C., Pallier, C., LeBihan, D., Dehaene, S., Dupoux, E. (2003). Phonological grammar shapes the auditory cortex: a functional magnetic resonance imaging study. *Journal of Neuroscience. 23*, 95419546.

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and toward logit mixed models. *Journal of Memory and Language, 29*, 434-446.

Jiang, J., Auer, E. T. Jr., Alwan, A., Keating, P. A., & Bernstein, L. E. (2007). Similarity structure in visual speech perception and optical phonetic signals. *Perception & Psychophysics, 69*, 1070-1083.

Jun, J. (1995). Perceptual and articulatory factors in place assimilation : an Optimality-theoretic approach. PhD dissertation, University of California, Los Angeles.

Jun, J. (1996). Place assimilation is not the result of gestural overlap: evidence from Korean and English. *Phonology, 13*, 377-407.

Jun, J. (2000). Preliquid nasalization.*Korean Journal of Linguistics, 25*(2). 191-208.

Jun, S. -A. (1994). The status of the lenis stop voicing rule in Korean. In Kim-Renaud, Y.-K. (ed.), *Theoretical Issues in Korean Linguistics* (pp. 101-114). CSLI.

Jusczyk, P. W., & Aslin, R. N. (1995). Infants' detection of sound patterns of words in fluent speech. *Cognitive Psychology, 29*, 1-23.

Jusczyk, P. W., Cutler, A., & Redanz, N. (1993). Preference for the predominant stress patterns of English words. *Child Development, 64*, 675-687.

Jusczyk, P. W., Houston, D. M., & Newsome, M. (1999). The beginning of word segmentation in English-learning infants. *Cognitive Psychology, 39*, 159-207.

Kabak, B., & Idsardi, W. (2007). Perceptual Distortions in the Adaptation of English Consonant Clusters: Syllable Structure or Consonantal Contact Constraints? *Language and Speech, 20*(1), 23-52.

Kahn, D. (1976). Syllable-based generalizations in English phonology. PhD dissertation. MIT.

Kawahara, S. (2015). The phonetics of sokuon, or obstruent geminates. In H. Kubonozo (ed.), *The Handbook of Japanese Language and Linguistics: Phonetics and Phonology* (pp. 43-73). Mouton.

Kawahara, S., & Garvey, K. (2014). Nasal place assimilation and the perceptibility of place contrasts. *Open Linguistics, 1*, 17-36.

Kiesling, S., Dilley, L., & Raymond, W. (2006). The Variation in Conversation (ViC) Project: Creation of the Buckeye Corpus of conversational speech. Department of Psychology, Ohio State University, Columbus, OH, available at www.buckeyecorpus.osu.edu (Last viewed 11/08/2015).

Kim, J., Davis, C., & Cutler, A. (2008). Perceptual tests of rhythmic similarity: II. Syllable rhythm. *Language and Speech*, 51, 343-359.

Kessler, B., & Treiman, R. (1997). Syllable structure and the distribution of phonemes in English syllables. *Journal of Memory and Language, 37*, 295-311.

Kim-Renaud, Y.-K. (1974). Korean Consonatal Phonology. Ph.D dissertation. University of Hawaii. Honolulu.

Krishnamurthy, R. (2000). Size matters: Creating dictionaries from the worlds largest corpus. In *Proceedings of KOTESOL 2000: Casting the Net: Diversity in Language Learning*, Taegu, Korea, 169180.

Kochetov, A., & Pouplier, M. (2008). Phonetic variability and grammatical knowledge: an articulatory study of Korean place assimilation.*Phonology, 25*, 399-431.

Kochetov, A., & So., C. K. (2007). Place assimilation and phonetic grounding: A cross-linguistic perceptual study. *Phonology, 24*(3), 397-432.

Kochetov, A., Pouplier, M., & Son, M. (2007). Cross-language differences in overlap and assimilation patterns in Korean and Russian. *Proceedings of the 16th International Congress of Phonetic Sciences*. 1361-1364.

Kohler K. (1990). Segmental reduction in connected speech: Phonological facts and phonetic explanations. In Hardcastle, W.J. & A. Marchal (eds.), *Speech Production and Speech Modeling* (pp. 69-92). Dordrecht: Kluwer Academic Publishers.

Kubozono, H. (2015a). Introduction to Japanese phonetics and phonology, In Haruo Kubozono (ed.) *The Handbook of Japanese Language and Linguistics: Phonetics and Phonology* (pp. 1-42) Mouton De Gruyter.

Kubozono, H. (2015b). Loanword phonology, In Haruo Kubozono (ed.) *The Handbook of Japanese Language and Linguistics: Phonetics and Phonology* (pp. 313-362) Mouton De Gruyter.

Kubozono, H., Itô, J., & Mester, A. (2008). Consonant gemination in Japanese loanword phonology. In *Current issues in unity and diversity of languages. Collection of papers selected from the 18th International Congress of Linguists*, ed. The Linguistic Society of Korea, 953973. Republic of Korea: Dongam Publishing Co.

Kuhl, P., (2004). Early language acquisition: cracking the speech code. *Nature Reviews Neuroscience, 5*, 831-843.

Ladefoged, P., & Johnson, K. (2015). *A Course in Phonetics* (7th edition). Stamford: Cengage. Learning Lee, H. (2002) . *Korean Standard Pronunciation Dictionary.* (in Korean) Seoul: Seoul National University Press.

Lee, Y. (2006). Sub-syllabic constituency in Korean and English. PhD dissertation, Northwestern University.

Lee, Y., & Goldrick, M. (2008). The emergence of sub-syllabic representations. *Journal of Memory and Language, 59*(2), 155-168.

Lee, Y., & Goldrick, M. (2011) The role of abstraction in constructing phonological structure. presented at 2011 LSA Annual Meeting.

Liberman A. M., Cooper F. S., Shankweiler D. P., & Studdert-Kennedy M. (1967). Perception of the speech code. *Psychological Review 74*, 431-461.

Malécot, A. (1956). Acoustic cues for nasal consonants: An experimental study involving a tape-splicing technique, *Language, 32*, 274-284.

Martinet, A. (1952). Function, structure, and sound change. *Word, 8*, 1-32.

Massaro, D. W. & Oden, G. C. (1980). Speech perception: A framework for research and theory. In N.J. Lass (Ed.), *Speech and Language: Advances in Basic Research and Practice. Vol. 3*, New York: Academic Press, 129-165.

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods, 44*(2), 314-324.

Mathesius, V. (1931). Zum Problem der Belastungs- und kombinationsfhigkeit der Phoneme. *Travaux du Cercle Linguistique de Prague, 4*, 148-152.

Mattys, S. L., & Melhorn, J. F. (2005). How do syllables contribute to the perception of spoken English? insight from the migration paradigm. *Language and speech, 48*(2), 223-253.

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

McNeil, D., & Lindig, K. (1973). The perceptual reality of phonemes, syllables, words, and sentences. *Journal of Verbal Learning and Verbal Behavior, 12*(4), 419-430.

Mehler, J., Dommergues, J., Frauenfelder, U., & Segui, J. (1981). The syllable's role in speech segmentation. *Journal of Verbal Learning, 20*, 298-305.

Meltzoff, A. N. & Moore, M. K. (1977). Imitation of facial and manual expressions by human neonates. *Science, 198*, 75-78.

Miller, G. A., & Nicely, P., (1955). An analysis of perceptual confusions among some English consonants. *The Journal of the Acoustical Society of America, 27*(2), 338-352.

Mills. A. B. (1983). Acquisition of speech sounds of the visually handicapped child. In A.E. Mills (ed.), *Language acquisition in the blind child: Normal and deficient* (pp. 46-56). London et al.: Croom Helm.

Mohr, B. & Wang, W. S. Y. (1968) Perceptual distance and the specification of phonological features.*Phonetica, 18*. 31-45.

Nazzi, T., Iakimova, G., Bertoncini, J., Fredonie, S., & Alcantara, C. (2006). Early segmentation of fluent speech by infants acquiring French: Emerging evidence for crosslinguistic differences. *Journal of Memory and Language, 54*(3), 283-299.

Norris, D., & Cutler, A. (1988). The relative accessibility of phonemes and syllables. *Perception & Psychophysics, 43*(6), 541-550.

Oh, M. (1994). A Reanalysis of consonant simplification and s-neutralization. In Kim-Renaud, Y.-K. (ed.), *Theoretical Issues in Korean Linguistics* (pp. 157-174). CSLI.

Oh, Y., Pellegrino, F., Coupé, C., & Marsico, E. (2013). Cross-language Comparison of Functional Load for Vowels, Consonants, and Tones, *Proceedings of Interspeech*.

Ohala, J. J. (1981). The listener as a source of sound change. In C.S. Masek, R. A. Hendrick, and M. F. Miller (eds.), *Para-session on Language and Behavior* (pp. 178-203), Chicago: Chicago Linguistic Society.

Pellegrino, F., Coupé, C., & Marisco, E. (2011). Across-language perspective on speech information rate. *Language, 87*, 539-558.

Pintér, G. (2015). The emergence of new consonant contrasts. In Haruo Kubozono (ed.) *The Handbook of Japanese Language and Linguistics: Phonetics and Phonology* (pp. 121-166) Mouton De Gruyter.

Pitt, M. A., Dilley, L., Johnson, K., Kiesling, S., Raymond, W., Hume, E. (2007). Buckeye Corpus of Conversational Speech (2nd release) [www.buckeyecorpus.osu.edu]

Polka, L., & Sundara, M. (2012). Word segmentation in monolingual infants acquiring Canadian English and Canadian French: Native language, cross-dialect, and cross-language comparisons. *Infancy, 17*(2), 198-232.

Pols, L. (1983). Three mode principle component analysis of confusion matrices, based on the identification of Dutch consonants, under various conditions of noise and reverberation. *Speech Communication, 2*, 275-293.

Raymond, W. D., Dautricourt, R., & Hume, E. (2006). Word-medial /t,d/ deletion in spontaneous speech: Modeling the effects of extra-linguistic, lexical, and phonological factors. *Language Variation and Change, 18*, 55-97.

Repp, B. H. (1984). Closure duration and release burst amplitude cues to stop consonant manner and place of articulation. *Language and speech, 27*(3), 245-254.

Savin, H., & Bever, T. (1970). The non-perceptual reality of the phoneme. *Journal of Verbal Learning and Verbal Behavior, 9*, 295-302.

Sebastian-Galles, N. (2005). Cross-language speech perception. In D. B. Pisoni & R.E.Remez (eds.), *The Handbook of Speech Perception* (pp. 546-566). Oxford, UK: Wiley-Blackwell Publishing.

Segui, J., Frauenfelder, U., & Halle, P. (2001). Phonotactic constraints shape speech perception: Implications for sublexical and lexical processing. In E. Dupoux (ed.), *Language, Brain, and Cognitive Development* (pp.195-208). Cambridge, MA: MIT Press.

Segui, J., Frauenfelder, u., & Mehler, J. (1981). Phoneme monitoring, syllable monitoring and lexical access. *British Journal of Psychology, 72*, 471-477.

Seyfarth, S., (2014). Word informativity influences acoustic duration: effects of contextual predictability on lexical representation. *Cognition, 133*, 140-55.

Shannon, C. E., (1948). A mathematical theory of communication.*Bell System Technical Journal, 27*(3-4), 379-423. 623-656.

Shannon, C. E., & Weaver, W. (1949). *The Mathematical Theory of Communication.* Urbana: University of Illinois Press.

Shibatani, M. (1990). *The Languages of Japan.* Cambridge: Cambridge Universigy Press.

Singh, S. & Black, J. (1966) Study of twenty-six intervocalic consonants as spoken and recognized by four language groups.*The Journal of the Acoustical Society of America, 39*(2), 372-387.

So, C. K., & Best, C. T. (2010). Cross-language Perception of Non-native Tonal Contrasts: Effects of Native Phonological and Phonetic Influences. *Language and Speech, 53*(Pt 2), 273293.

Sohn, H.-M. (1999). *The Korean Language.* Cambridge: Cambridge University Press.

Son, M., Kochetov, A., & Pouplier, M. (2007). The role of gestural overlap in perceptual place assimilation: evidence from Korean. In Jennifer Cole & Jose Ignacio Hualde (eds.), *Laboratory Phonology, 9.* Berlin & New York: Mouton de Gruyter.

Stemberger, J. P. (1991). Apparent anti-frequency effects in language production: The addition bias and phonological underspecification. *Journal of Memory and Language, 30*, 161-185.

Sumby, W., & Pollack, I. (1954). Visual Contribution to Speech Intelligibility in Noise.*The Journal of the Acoustical Society of America, 26*, 212-215.

Surendran, D., & Niyogi, P. (2003). Measuring the usefulness (functional load) of phonological contrasts, Technical Report TR-2003-12, Dept. of Comp. Science, Univ. of Chicago.

Surendran, D., & Niyogi, P. (2006). Quantifying the Functional Load of Phonemic Oppositions, Distinctive Features, and Suprasegmentals. In O. N. Thomsen (ed.), *Competing*

Models of Language Change: Evolution and Beyond (pp. 4358). Amsterdam; Philadelphia, PA: John Benjamins.

The Sejong Spoken Corpus (2009). The National Institute of the Korean Language. [CD-ROM]

Tsujimura, N. (2013). *An Introduction to Japanese Linguistics*. 3nd edn. Wiley-Blackwell.

Vance, T. J. (2008). *The sounds of Japanese*. Cambridge: Cambridge University Press.

Wang, W. S. Y., & Fillmore, C.,J. (1961). Intrinsic cues and consonant perception. *Journal of Speech and Hearing Research, 4*, 130-136

Wedel, A., Kaplan, A., & Jackson, S. (2013). High functional load inhibits phonological contrast loss: a corpus study. *Cognition, 128*, 179-86.

Weide, R. L. (1994). CMU Pronouncing Dictionary. http://www.speech.cs.cmu.edu/cgi-bin/cmudict.

Winters, S. (2001). VCCV perception: Putting place in its place. In E. Hume, N. Smith & J. van de Weijer (eds.), *Surface Syllable Structure and Segment Sequencing* (pp. 230-247), Leiden, NL: Holland Institute of Linguistics.

Winters, S. (2003). Empirical investigations into the perceptual and articulatory origins of cross-linguistic asymmetries in place assimilation. PhD dissertation, Ohio State University.

Yoneyama, K. (2000). Phonological neighborhoods and phonetic similarity in Japanese word recognition. PhD dissertation, Ohio State University.

Zipf, G. (1929). Relative frequency as a determinant of phonetic change. *Harvard Studies in Classical Philology, 40*, 1-95.

Zipf, G. K. (1968). *The Psycho-biology of Language. An Introduction to Dynamic Philology.* Cambridge, MA: The MIT Press.