Japanese Co-Occurrence Restrictions Influence Second Language Perception

Alexander J. Kilpatrick

University of Melbourne

Rikke L. Bundgaard-Nielsen

MARCS Institute for Brain Behaviour and Development, Western Sydney University

Brett J. Baker

University of Melbourne

Abstract

Most current models of L2 speech perception (e.g., PAM-L2; Best & Tyler, 2007; SLM; Flege, 1995; NLM; Kuhl, 1993) base their predictions on the native/non-native status of individual phonetic/phonological segments. This paper demonstrates that the phonotactic properties of Japanese influence the perception of natively contrasting consonants and suggests that phonotactic influence must be formally incorporated in these models. We first propose that by extending the perceptual categories outlined in PAM-L2 to incorporate sequences of sounds, we can account for the effects of differences in native and non-native phonotactics on non-native and cross-language segmental perception. Secondly, we test predictions based on such an extension in two perceptual experiments. In Experiment 1, Japanese listeners categorised and rated /VCV/ strings in combinations that either obeyed or violated Japanese phonotactics. The participants categorised phonotactically illegal strings to the perceptually nearest (legal) categories. In Experiment 2, participants discriminated the same strings in AXB discrimination tests. Our results show that Japanese listeners are more accurate and have faster response times when discriminating between legal strings than between legal and illegal strings. These findings expose serious shortcomings in currently accepted L2 perception models which offer no framework for the influence of L1 phonotactics.

Keywords: Phonology, Japanese, Phonotactics, Second Language Perception, Perceptual Assimilation.

1. Introduction

Current non-native (L2) perception models posit that the phonological (abstract categories of 'sound' in a language) and phonetic (acoustic/articulatory realisations of the phonemic categories) contrasts of a speaker's native language (L1) systematically influence how learners perceive phones in a second language (L2) (Best, 1995; Best & Strange, 1992; Best & Tyler, 2007). Such models (e.g., PAM-L2; Best & Tyler, 2007; SLM; Flege, 1995; NLM; Kuhl, 1993, see discussion below) have primarily focused on the effect that non-overlapping and/or partially overlapping phonemic and phonetic categories have on L2 speech perception without formal consideration of the role of phonotactics in L2 perception. Existing research has typically focussed on situations in which L2 listeners perceptually assimilate two or more L2 phones into a single native phonemic category; this situation commonly results in difficulties discriminating the two phones in question. The degree of difficulty experienced by listeners in such studies is often quantified in terms of consistency of L2-to-L1 category assimilation patterns and L2 contrast discrimination accuracy. The degree of difficulty of these tasks has, however, also been found to increase the response time (RT) of participants, and this increase in RT is generally assumed to be a manifestation of increases in cognitive load (e.g., Adank et al., 2009; Munro & Derwing, 1995; Schmid & Yeni-Komshian, 1999).

A number of existing studies have shown that L1 phonotactics influence L2 perception (e.g., Dupoux et al., 2011; Halle, et al., 2008; Kabak & Idsardi, 2007). These studies have typically found that listeners sometimes misperceive speech sequences that violate native phonotactics as those that do not. However, this research has almost exclusively focused on constraints on consonant sequences, such as the consonant clusters in /fmatu/ or /tlabod/ (e.g., Davidson & Shaw, 2012; Hallé, et al., 1998). Limited research to date has examined the role of co-occurrence restrictions at the consonant-vowel (CV) level, despite results from 40 years of phonological research suggesting that the syllable plays a pervasive role in shaping the sounds of languages (e.g., Blevins, 1995; Kahn, 1976). The small volume, and narrow scope, of experimental work examining the phonotactic shaping of perception likely limits current theories of speech perception by disregarding phonotactic constraints on the distribution of L1 phones and the role that such constraints may also have in L2 perception.

The present paper addresses this gap and demonstrates that phonotactic properties—beyond the constraints on consonant clusters—influence segmental perception, in this case the perception of English consonants by Japanese learners of English. Through two experiments,

we show that Japanese co-occurrence restrictions—which limit how Japanese consonants and vowels may combine in CV sequences—systematically influence L2 perception. The first experiment examines how Japanese listeners categorise /VCV/ sequences (hereafter, "strings") of phones that occur in Japanese which either violate (hereafter, "illegal strings") or adhere to (hereafter, "legal strings") Japanese phonotactic constraints (see Table 1). The results show that illegal strings are categorised as instances of the perceptually nearest legal category but, on average, are assigned a lower goodness-of-fit rating than their legal counterparts, suggesting that listeners typically perceive illegal strings as less canonical than legal strings.

The second experiment examines the discriminability of legal and illegal strings in a series of AXB discrimination tests (Table 1). The results indicate that Japanese phonotactic constraints systematically shape segmental perception, such that even contrastive L1-occurring phonemes become difficult to discriminate when one phone occurs as part of an illegal sequence. Importantly, these results show that the relationship between a phoneme and its phonetic realisation is dependent on the sequence in which the phone occurs and that sequence's adherence to the phonotactic constraints of the L1. We account for these findings by extending one of the prevailing models of speech perception, the Perceptual Assimilation Model (PAM/PAM-L2; Best, 1995; Best & Tyler, 2007). However, PAM-L2 only considers the effects of non-overlapping phonetic and phonemic inventories and would therefore predict that contrasting legal strings such as /uʃu/-/usu/ would be equally discriminable to contrasting strings that vary in their legality, such as /iʃi/-/isi/. The experiments detailed in the present paper show that this is not the case, and we therefore propose that an extension to PAM-L2 is necessary to predict and account for this variance. This extension is discussed in Section 1.3. In what follows, we firstly discuss the crucial aspects of Japanese phonotactics that are required to understand the foundations of the experimental sections which follow.

1.1 Japanese phonotactics

Japanese phonology is prominently characterised by its restrictive phonotactic system, which disallows non-homorganic consonant clusters (Tsujimura, 2013). The limits on Japanese consonant sequences have indeed been the focus of most studies that have examined the influence of phonotactically conditioned constraints on perception, (e.g., Dupoux et al., 1999; 2011). In addition to the constraints on consonant sequences, Japanese also has numerous co-occurrence restrictions on possible /CV/ combinations (Kawahara et al., 2006; Tsujimura, 2013), and we argue that these constraints provide a unique opportunity to examine the

interface between phonology and phonetics. Indeed, Japanese phonotactic constraints on /CV/ combinations allow us to show that the typical association between a phoneme and its phonetic realisation such as between /s/ and [s] can be disrupted by phonotactic constraints to such an extent that [s] will be more readily assimilated to /ʃ/ than to /s/ under certain conditions, despite the fact that /s/ and /ʃ/ are contrastive phonemes in the language. The results suggest that such re-mappings influence both discrimination accuracy and the time it takes for participants to respond to discrimination tests.

It is likely that such perceptual behaviour is linked to the observation that contemporary Japanese speakers fall along a continuum which determines how closely they adhere to Japanese phonotactics when producing loanwords with phonotactically illegal segmental combinations (Pintér, 2015). Indeed, it is probable that the introduction of recent loanwords has led to a 'split' in some of these phonemes so that erstwhile allophones have now become contrastive in environments in which the contrast was formerly predictable (Mutsukawa, 2009; Pintér, 2015; Vance, 1987: Possible innovative CV combinations are shown in Table 2). To our knowledge, the present studies provide the first experimental support for such a language shift with associated changes in perception.

Itô and Mester (1999) propose that Japanese phonological rules, including phonotactic restrictions, exist in a hierarchy which reflects the stratified structure of the lexicon. They maintain that lexical items in the more established strata adhere to a greater number of these phonological restrictions. The rules (expressed as phonological constraints in an Optimality Theoretic model of the grammar) provide a framework for a 'Core-Periphery' structure in which the distance from the core indicates the degree of integration of lexical items (Pintér, 2015): items that are closer to the core are more native-like while items that are on the periphery are less native-like. The concept of constraint hierarchies can be applied directly in the analysis of weakening phonotactics: innovative /CV/ combinations can be accounted for by extending the lexicon with new constraints at the periphery. For example, Conservative Japanese maintains the constraint */si/ and the constraint */ti/ so one might expect that an English word like ‘city’ /siti/ would be borrowed by Conservative Japanese speakers as /ʃitʃi/. Itô and Mester propose that these constraints exist at unequal distances from the core of their model, with */ti/ occurring further from the centre than */si/, meaning that the constraint against /si/ sequences is more strongly maintained than the constraint against /ti/ sequences, for all speakers. They argue that one might expect an Innovative dialect speaker to produce the word ‘city’ as /ʃitʃi/,

/ʃiti/ or /siti/; however, /sitʃi/ contravenes the faithfulness rankings in the loanword stratum. The Core-Periphery model finds a direct reflection in the perceptual behaviour of the speakers in this study, thus supporting Itô & Mester's proposals.

In the following, we propose that the difference in ranking between the */si/, and */ti/ constraints likely influences the rate at which listeners are exposed to each sequence in Japanese discourse and argue that this results in perceptual tuning, allowing for a more accurate disambiguation between predictable sequences and their perceptually nearest legal counterparts (/ʃi/, and /tʃi/ respectively). We propose a similar relationship between /ti/ and /tu/. Tamaoka and Makioka (2004) have conducted a corpus analysis of over half a billion CV sequences in Japanese newspapers which includes most innovative sequences. While they provide counts of /ti/ (66,313 occurrences) and /tu/ (1,854 occurrences), they do not provide a /si/ count. This is likely because their interpretation of the kana (script) is based upon a report by the Cabinet of Japan's Ministry of Education, Culture, Sports, Science and Technology (1991) which offers no standardised method for transcribing /si/ (see Table 3). Therefore, we can predict that /ti/ likely occurs more frequently than /si/ and /tu/ in spoken Japanese discourse, but we can make no such prediction as to the relative frequency of /si/ and /tu/.

1.2 Segmental perception and phonotactics

A number of models of speech perception, including the Speech Learning Model (SLM; Flege, 1995), the Native Language Magnet Model and its expansion (NLM/NLM-e; Kuhl, 1993; Kuhl et al., 2008) and the Perceptual Assimilation Model and its extension to L2 perception (PAM/PAM-L2; Best, 1995; Best & Tyler, 2007), aim to account for the well-documented difficulties experienced by L2 language learners and participants in cross-language and L2 perception experiments. PAM-L2, developed by Best and colleagues (Best, 1994, 1995; Best & Tyler, 2007), contends that non-native perception is predictable if the degree of articulatory-phonetic or gesture similarity between native and non-native phones is taken into account. PAM-L2 proposes that non-native speech perception is robustly filtered by listeners' linguistic backgrounds, so that listeners assimilate non-native sounds into their nearest phonemic (L1) category. This accounts for the well-documented difficulty Japanese listeners experience in distinguishing English /r/ from /l/. PAM-L2 asserts that these non-native sounds are perceptually similar due to their assimilation into a single Japanese category. The perceptual 'filter' is the result of what PAM-L2 describes as a process of perceptual 'tuning': the process whereby an individual's linguistic perceptual system adapts to focus on the

meaningful information contained within the phonological and phonetic systems of their native language (Kuhl et al., 1992; Werker & Tees, 1984). PAM-L2 also assumes that perceptual tuning and the acquisition of new phonetic categories is possible throughout the lifespan, although infants, children and adults perceive non-native speech differently from native speech because their perceptual systems have been tuned to their native language to varying degrees as a function of native language exposure.

Research also suggests that perceptual tuning involves two processes: perceptual broadening and perceptual narrowing (Scott et al., 2007), which simultaneously improve the ability to detect and differentiate between commonly occurring stimuli, while causing a decline in the ability to differentiate between uncommon stimuli (e.g., Lewkowicz, 2014; Scott et al., 2007; Scott & Monesson, 2010). Perceptual narrowing is demonstrated by studies that examine the 'untuned' perceptual systems of infants, which find that infants can discriminate between unfamiliar, non-native phones more accurately than adults from the same speech community (e.g., Werker et al., 1981; Werker et al., 2007). Once the perception system has been 'tuned' (in adulthood), native speakers experience varying degrees of difficulty in perceiving speech sounds which do not represent native contrasts for them. PAM-L2 suggests a number of ways in which non-native segments may be perceived in relation to native phones; these include, in gradient fashion: non-native phones which may be perceived as a good example of a native phonemic category; an adequate example of a native phonemic category; a deviant example of a native phonemic category; and a phone that falls completely outside of any phonemic category.

On the basis of these patterns, PAM-L2 predicts the extent of discriminability of a given non-native contrast:

1) 'Two Category Assimilation' (TC), where both non-native phones are perceived as instances of two native phonemic categories. TC distinctions are expected to be highly discriminable because TC pairs are mapped to different native phonemic categories.

2) 'Category Goodness Difference' (CG), where both non-native phones are perceived as members of a single native phonemic category, but differ in the goodness-of-fit to that particular native category. Discrimination of CG contrasts is expected to range from medium to good depending on the magnitude of difference in the goodness-of-fit to the native phoneme category into which they are assimilated.

3) 'Single Category Assimilation' (SC), where both non-native phones are perceived as equally good instances of a single native phonemic category. SC distinctions are expected to be difficult to distinguish, because both phones map to a single category equally well.

4) 'Both Uncategorisable' (UU), where both non-native phones fall in-between two or more existing native phonemic categories. PAM-L2 considers both phonetic and phonological elements when explaining the discriminability of phones that have no resemblance to native categories. UU distinctions are predicted to range from poor to moderate discriminability, depending on their perceptual proximity to each other and existing categories.

5) 'Uncategorisable versus Categorisable' (UC), where one non-native phone falls between two or more native categories while the other is perceived as an instance of a native category. UC distinctions are often easily distinguished because one sound is mapped to an existing phonemic category, while the other is not recognisable and is therefore perceived as different to the mapped phone.

6) 'Non-assimilable' (NA), where non-native speech sounds are so divergent from native speech sounds that they are not processed as speech. In these instances, phone discriminability may range from good to excellent, depending on the perceptual difference between the divergent phones.

While PAM-L2 provides predictions for the ways in which individuals will perceive non-native segments, these predictions are made exclusively on the basis of differences in the phonemic inventories of two languages, and the phonetic realisations of these phones. PAM-L2 does not take into account the influence of phonotactics, despite findings suggesting that native phonotactics systematically influence non-native segmental perception in some cases (e.g., Cuetos et al., 2011; Dupoux, et al., 1999; Hallé & Best, 2007; Hallé, et al., 1998; Wagner, et al., 2012). For instance, researchers have observed that French listeners perceptually repair word initial /tl/ and /dl/ clusters as /kl/ and /gl/ (Hallé, et al., 1998), likely because word-initial /tl/ and /dl/ clusters are a violation of French phonotactics, whereas clusters of velar stop plus lateral are phonotactically legal. A limited number of previous studies have, however, relied on PAM to provide a theoretical framework to explain the influence of phonotactics on non-native/L2 segmental perception. For example, in a study of Japanese perception of non-native phonology, it is argued that Japanese listeners perceptually repair consonant clusters with

epenthetic vowels, as in the case of /ebzo/→/ebuzo/ (Dupoux et al., 1999). Here, the researchers propose that PAM might be altered to span larger chunks of signal (Dupoux et al., 1999). Following this suggestion, we present such an extension of PAM to the level of phonotactics, which proposes that monolingual Japanese speakers decode the speech signal as individual units but are tuned to the frequency at which units co-occur.

1.3 Extending PAM-L2

PAM-L2 argues that listeners are tuned to the phonological and phonetic contrasts of their L1 and in the following, we extend this model to incorporate the influence of L1 phonotactics in cross- and second language perception. We propose that listeners perceive and map cross-language and second language input not only in terms of individual phones (which has generally been the focus in PAM/PAM-L2 based research) but also in terms of the transitional probability of sequences of sounds. Phonotactically prohibited sequences of phones are mapped to sequences that are both perceptually similar to the incoming signal and are transitionally probable. In the context of the present paper, this extension predicts that stimuli which contain diphones that Japanese listeners are infrequently exposed to in their L1 (such as /si/, /tu/ and /ti/) will sometimes be assimilated to their perceptually nearest legal match (/ʃi/, /tsu/ and /tʃi/). Consistent with PAM-L2, we propose that listeners are sensitive to the frequency at which sounds co-occur and because these phonotactically illegal diphones vary in how frequently they occur in Japanese discourse, we predict that illegal tokens will also vary in their discriminability against their legal matches. Here, we predict that the /itʃi/-/iti/ contrast will be easier to distinguish than /iʃi/-/isi/ and /utsu/-/utu/ contrasts due to the likely increased frequency with which /ti/ occurs in Japanese discourse. This interpretation would suggest that frequently occurring sequences behave as instances of existing native categories, while less frequently occurring sequences behave as less canonical instances of their perceptually nearest match, depending on how often listeners are exposed to them. Essentially, our model proposes that certain patterns of segment assimilation outlined in PAM-L2 can be extended to sequences of speech, and while PAM-L2 only considers contrasting phonetic and phonemic inventories, our extension also considers the frequency at listeners are exposed to certain sequences of sounds.

We propose that, by extending PAM-L2 to incorporate phonotactic constraints on non-native phoneme sequences, the above /ebzo/-/ebuzo/ example can be framed as a PAM-L2 CG contrast, whereby the illegal consonant cluster is undergoing perceptual assimilation. Due to

restrictions in Japanese phonotactics, Japanese listeners are infrequently exposed to non-homorganic consonant clusters and /VbzV/ is thereby sometimes assimilated to its most perceptually similar match, /VbuzV/. It is important to note that in this case, /u/ is epenthesised because the Japanese /u/ is the most phonetically minimal Japanese vowel (Dupoux et al., 2011), making it a nearer match to the missing sound.

In the light of our extension of PAM-L2 presented above, we can formulate a number of general predictions for how native phonotactic constraints shape the perception of non-native phonological (and phonetic) segments. In general, we predict that non-native listeners with limited L2 experience (i.e. listeners whose phonology is closer to the 'Core' in Itô & Mester's 1995 terms) will more accurately discriminate between sequences of phones which are assimilated into two discrete native phonotactic structures (TC contrasts) and less accurate at discriminating between sequences of phones which are assimilated into the same native phonotactic structure, whether equally well (an SC contrast) or with differences in the goodness-of-fit to the native category (a CG contrast). Just as is generally predicted for SC contrasts, it is likely that non-native phonotactic strings that are assimilated equally well into the same native phonotactic structure will be near-impossible to discriminate. We also predict that TC contrasts will require less processing time than CG contrasts, resulting in differences in response time on perception tasks involving these contrasts. In summary, we predict that tests that measure the discriminability of phonotactically illegal and legal di-phones are more difficult than those that measure between two legal di-phones, and that this will increase both inaccuracy and response time.

In the following, we present a two-part experiment—Experiment 1: Categorisation (Section 2) and Experiment 2: Discrimination (Section 3)—investigating the perception of English /VCV/ strings, some of which violate Japanese phonotactic constraints, by Japanese learners of English. Experiment 1 is a cross-language mapping task in which Japanese listeners categorise non-native strings (legal and illegal) into existing Japanese categories and assign goodness-of-fit ratings to indicate how well these strings 'fit' the Japanese categories. Here, we predict that participants will categorise illegal strings, /isi/, /iti/ and /utu/ into their perceptually nearest legal category, respectively, /iʃi/, /itʃi/ and /utsu/. We predict that participants will assign phonotactically illegal strings a lower goodness-of-fit rating than legal strings. In Experiment 2 the same Japanese participants discriminate a series of contrastive strings consisting of pairs of legal strings as well as pairs of legal and illegal strings. We also

predict that discrimination of two legal strings will be easy (corresponding to a TC phonemic contrast) because chunks within both strings will be mapped to existing sequences. We also predict that legal-illegal string contrasts will be less accurately discriminated, corresponding to PAM-L2 CG phonemic contrasts, and that this drop in accuracy will be accompanied by an increase in response time.

2. Experiment 1: Categorisation and Goodness-of-fit

2.1 Method

2.1.1 Stimuli

The stimuli were produced by three female Australian English (AustE) speakers, who were PhD candidates in the field of linguistics and have phonetic training. The recordings took place in the Horwood recording studio located at the University of Melbourne, using a Charter Oak E700 condenser microphone and Tieline Commander G3 codec protocol recording in mono with a bit depth of 64kb/sec and a sample rate of 48kHz. Speakers were asked to produce strings in accordance with AustE stress patterns. Each speaker produced five consecutive repetitions of eight vowel-consonant-vowel (/VCV/) strings, and strings 2, 3, and 4 were selected as stimuli (to avoid tokens with potential hesitation due to list initial unfamiliarity and tokens with list final intonation). The /VCV/ strings were excised using a *Praat* (Boesrsma & Weenink, 2015) script and manually checked. Using a second Praat script, the excised strings were normalised for volume and enveloped with a 20 millisecond ramp-in and a 10 millisecond ramp-out.

The stimuli included five /VCV/ combinations that adhere to Conservative Japanese co-occurrence restrictions; /iʃi/, /itʃi/, /usu/, /uʃu/ and /utsu/, and three that do not; /isi/, /iti/ and /utu/. Speakers were asked specifically to produce /itʃi/ and /utsu/ strings as affricates. All but one (/uʃu/) of the phonotactically legal strings used in our experiments constitute existing Japanese lexical items: /itʃi/ means 'one'; /iʃi/ 'stone'; /utsu/ 'depression'; and /usu/ ‘mill-stone’. This is an inevitable outcome of a language with a very limited phonological system, in which it is difficult to randomly combine legal sequences without creating existing Japanese words.

2.1.2 Participants

Twelve female native speakers of Japanese participated in this study. All participants were recruited by word of mouth. Participant ages ranged from 27 to 34 (M = 29.6, SD = 2.22). All participants were born in the Yamagata prefecture of Japan and had spent the majority of their lives in Yamagata. Some of the participants had holidayed in a foreign country, however none of them had spent more than one month in total outside of Japan. All participants had undertaken six years of compulsory English instruction through middle and senior high school. Depending on the school, these classes occasionally (ranging from weekly to monthly)

involved additional instruction from native English speaking assistant language teachers (ALTs). Other than these occasional ALTs, the participants' English instruction in compulsory classes were conducted by native Japanese teachers.

Five of the participants had undertaken further English instruction outside of the compulsory six years, ranging from half a year to four years of further study (M = 1.58, SD = 1.42). Of these five, only one participant had received instruction from a native English speaker. This instruction was conducted by a native British English speaker and occurred in an English language extended-hours school which the participant attended for four years. None of the participants had studied an additional language other than English. Two participants were excluded from the analyses: One participant's experiment was disrupted and she was unable to complete the tasks; the other failed to understand the task.

2.1.3 Procedure

All experiments were conducted in quiet rooms in Tokyo and Murayama City. The experiment was presented over noise cancelling headphones from a MacBook Pro using PsyScope X (Bonatti, 2015). Before the experiment commenced, all participants completed a detailed language background questionnaire and were given an instructional handout, both of which were presented in Japanese.

Experiment 1 consisted of a categorisation experiment in which the participants were exposed to randomised presentations of the selected /VCV/ strings and asked to categorise each string into one of four string categories (discussed in detail below). The experimental strings were drawn from a library consisting of nine repetitions, three from each speaker, of the eight /VCV/ string types listed in section 2.1.1. The four categorisation options were presented in hiragana orthography on 'buttons' on the computer screen, as shown in Table 4. Participants only had access to four categories, despite there being orthographical conventions for transcribing most of the phonotactically illegal strings. These options for categorising illegal strings were not provided in order to determine how Japanese speakers categorise strings according to Conservative Japanese. Participants were not given any information regarding the language in which the strings had been produced, however, given that experiments were conducted by an English-speaking non-Japanese researcher, participants likely assumed that strings would be produced by non-Japanese speakers.

Upon selecting one of the four available string categories, participants would be presented again with the same string and asked to assign a goodness-of-fit rating to the category match, along a scale from 1 to 7.

The categorisation trials were spaced with a 1500 millisecond inter-trial interval (ITI). There was a 1000 millisecond pause between categorising and goodness-of-fit rating tasks. If participants failed to respond to either task within a 3500 millisecond response window, the entire missed trial would be replayed at a random time during the remainder of the experiment. All data from tasks that timed out in this way were discarded. 200 categorisation and goodness-of-fit responses were elicited from each participant, 25 per /VCV/ string.

2.2 Predictions

Experiment 1 tests the following hypotheses:

H1) Strings that do not violate Japanese phonotactics will be accurately mapped to the nearest Japanese diphonic category. Hiragana is described as a phonetic orthographical system—and while this is not entirely accurate—Japanese categories were determined by the way in which hiragana graphemes typically correspond to International Phonetic Alphabet (IPA) symbols: い /i/, う /u/, し /ʃi/, ち /tʃi/, す /su/, and つ /tsu/. Therefore, we predict that /iʃi/ will be mapped to いし, /itʃi/ to いち, /usu/ to うす, and /utsu/ to うつ. We predict that participants will categorise /uʃu/, the only phonotactically legal string without a corresponding category presented as an on-screen option, as うす (/usu/) but allocate a low goodness-of-fit rating to reflect the perceptual distance between these native strings. Despite /uʃu/ being phonotactically legal, we did not provide an /uʃu/ category. We did this to provide identical experimental conditions between /iʃi/-/isi/ and /uʃu/-/usu/ discrimination trials in experiment 2.

H2) Strings that violate Japanese phonotactics will be mapped to the perceptually most similar Japanese category: /isi/ to いし (/iʃi/), /iti/ to いち (/itʃi/), and /utu/ to うつ (/utsu/).

H3) Japanese phonotactics will affect the goodness-of-fit ratings such that legal /VCV/ strings will achieve a higher goodness-of-fit rating than illegal strings because legal strings are perceptually closer to the selected Japanese category prototypes.

2.3 Categorisation Results

In the categorisation and goodness-of-fit experiment, participants largely categorised strings as predicted based on PAM-L2 and our extension to phonotactics, and in accordance with documented loanword assimilation patterns.

2.3.1 Categorisation

The categorisation performance of the Japanese participants was extremely consistent: All strings were assimilated into a single Japanese category at least 98% of the time (see Table 5 below). Having two strings, one which adheres to Japanese phonotactics and a perceptually similar string which does not, assimilate to a single native category suggests that listeners will perceive these strings as CG string contrasts. However, as mentioned above, we have not provided a category for the legal string /uʃu/ and, although both /uʃu/ and /usu/ were categorised within the /usu/ category, /ʃu/ and /su/ are contrastive in Japanese and are a likely TC contrast. We therefore argue that English /iʃi/-/iti/ and /uʃu/-/usu/ are perceived as likely TC contrasts (with predicted good discriminability) by Japanese listeners, while /itʃi/-/iti/, /iʃi/-/isi/, and /utsu/-/utu/ are perceived as likely CG contrasts (with predicted less accurate discriminability).

2.3.2 Goodness-of-fit ratings

As Table 5 (above) illustrates, the goodness-of-fit ratings indicate that participants perceived the two members of the three within-category (CG) /VCV/ contrasts differently, such that each of the three Japanese assimilation categories had a more prototypical member and a non-prototypical member. We infer this from the fact that there is a difference between the average goodness rating for those /VCV/ combinations that adhere to Japanese phonotactics (/iʃi/ = 5.96, /itʃi/ = 5.95, /usu/ = 5.59, /uʃu/ = 2.47, /utsu/ = 5.11) compared to those that violate co-occurrence restrictions (/isi/ = 4.23, /iti/ = 2.88, /utu/ = 2.71). Although /uʃu/ is a legal string, うしゅ/uʃu/ was not an available choice in the categorisation experiment and thus was categorised and ranked in accordance with the perceptually nearest available category, うす /usu/. Of the non-prototypical (illegal strings) /VCV/ goodness-of-fit ratings, the /isi/ string receives a rating which is surprisingly high (M = 4.23, SD = 0.76), compared to /uʃu/, (M = 2.47, SD = 0.44), /utu/, (M = 2.71, SD = 1) and /iti/, (M = 2.88, SD = 0.91).

To assess the significance of the differences in goodness-of-fit ratings between the more versus less canonical English /VCV/ strings assimilated into a single Japanese category, we

conducted a series of paired sample *t*-tests. As Table 6 illustrates, there was a significant difference between prototypical and non-prototypical strings in all tests. The perceptual distance between within-category strings can somewhat be inferred from the Cohen's *d* effect sizes in Table 6. These show that /utsu/-/utu/ strings were the most perceptually similar, while /uʃu/-/usu/ had the greatest perceptual distance.

As discussed above, the Japanese participants were extremely consistent in assimilating the English /VCV/ strings into the predicted Japanese categories, with no string assimilated less than 98% of the time (M = 99.25%, SD = 0.83). Indeed, categorisation consistency did not significantly differ for prototypical and non-prototypical strings, and we note that, of the 2000 categorisations included in these analyses, a total of only 15 strings were not categorised into expected categories. A detailed analysis of each of those 15 responses that were not consistent with our cross-language mapping predictions revealed that they typically received low goodness-ratings (M = 2.73, SD = 1.94).

2.4 Experiment 1: Discussion

The present experiment examines how near-monolingual Japanese listeners map non-native phoneme sequences that violate Japanese co-occurrence restrictions to phonotactically legal Japanese sequences. The results indicate that there is a prototypical and non-prototypical string mapped to each category which differ in goodness-of-fit ratings. This difference in goodness-of-fit rating suggests that those strings that are mapped to the same native category are CG distinctions. These results are consistent with our hypothesis (H2) that listeners categorise non-native strings into their perceptually closest category and in accordance with previous observations of the typical patterns of conservative loanword assimilation for /CV/ sequences. Indeed, no string was assimilated less than 98% of the time into the acoustically/articulatory most similar phonotactically legal Japanese /VCV/ string, and those few items that were not categorised as predicted received low goodness ratings, suggesting that participants were aware that the category match had poor goodness-of-fit.

The categorisation task results (Figure 1) also demonstrate—despite clear and highly consistent categorisation patterns for all English /VCV/ strings—that strings which violate Japanese phonotactics were given lower goodness-of-fit scores, likely reflecting their status as less acceptable or prototypical to native Japanese listeners. The series of paired sample *t*-tests calculated to test for significant differences between the goodness ratings of each string contrast

returned a significant difference between all prototypical and non-prototypical strings, clearly suggesting that strings were not perceived as equally good instances of assigned categories. A measurement of effect size (Cohen's *d*) calculated on the results of the paired sample *t*-tests (Table 6) indicated that those strings assimilated to the うす (/uʃu/-/usu/) category achieved the largest effect size of all categories. We propose that this difference in effect size is the result of a greater perceptual difference between the phonotactically legal and illegal /VCV/ strings because, as predicted by our extension of PAM-L2, these two particular strings (/uʃu/-/usu/) are being assimilated into two existing phonotactically acceptable Japanese strings. We hypothesise that, similarly to those processes outlined in PAM-L2, the perceptual distance, and in turn discriminability, of /usu/ and /uʃu/ is the result of these strings representing existing or legal Japanese strings. If this analysis is extended to the other native string categories, then the string pairs within these categories might be ranked in accordance with their perceptual distance. Under this analysis, the /uʃu/-/usu/ strings are the most perceptually dissimilar, followed in order by the /itʃi/-/iti/, /iʃi/-/isi/ and /utsu/-/utu/ strings, as reflected by the Cohen's d effect sizes presented in Table 6. This perceptual distance ranking of string pairs can be used to predict PAM-L2 assimilation types (Table 7), where TC test strings are predicted to be perceptually dissimilar while CG test strings are predicted to be perceptually more similar.

In Experiment 1, we assigned no assimilation criterion due to the extremely high assimilation rates as outlined in Table 5. Cross language mapping experiments that follow a similar methodology normally employ assimilation criteria of between 50-70% (e.g., Bundgaard-Nielsen et al., 2011). However, in the present paper, this was not necessary as no string achieved an assimilation rate of less than 98%. We propose that this high level of assimilation was due to both the relatively small number of categories within which strings might be assigned, and the perceptual distinctiveness of strings. Given that both the consonants and vowels differed across strings, participants were able to accurately assign strings to expected categories.

Finally, of the illegal tokens, /isi/ achieved the highest goodness-of-fit rating. Given that this finding is not consistent with the results presented in Experiment 2, we propose that this may be due to there being no standardised method for transcribing /si/. Indeed, we hypothesise that this lack of standardisation means that Japanese listeners are accustomed to /si/ being transcribed in multiple ways including in the katakana variant (シ) of the hiragana し (typically transcribed as /ʃi/) which was presented in the present experiment. Given that

categories were presented orthographically, it is possible that participants may have been more forgiving when assigning a score to this token due to its lack of standardisation.

3. Experiment 2: Discrimination

3.1 Method

3.1.1 Stimuli

The stimuli were the same /VCV/ strings used in Experiment 1. We categorised the five AXB discrimination contrasts into either TC categories or CG categories (see Table 7). The /iʃi/-/iti/ contrast was selected as a control whereby we could measure two legal tokens that are presumably easy to distinguish due to the medial consonants varying in their method of articulation. The /uʃu/-/usu/ contrast was selected to give an indication of the influence of the */si/ constraint when measured against the /iʃi/-/isi/ contrast. We predicted that the /uʃu/-/usu/ contrast was also relatively easy to discriminate but wanted to test the discriminability of two legal tokens that shared a phonetically similar medial consonant. The /itʃi/-/iti/ contrast provides a measure of the discriminability of a legal sequence against a sequence that is currently undergoing change in its status due to the weakening of the */ti/ constraint. The /iʃi/-/isi/ and /utsu/-/utu/ contrasts were selected to provide a measure of the discriminability of licit-illicit sequences.

3.1.2 Participants

The listeners were the same 10 native speakers of Japanese who successfully completed Experiment 1.

3.1.3 Procedure

Experiment 2 was conducted shortly after Experiment 1. The experimental conditions, including the experimental setting and the manner in which the task was explained to the participants, were identical to Experiment 1. Experiment 2 consisted of five separate AXB discrimination blocks (one per contrast), consisting of 60 trials, a total of 300 AXB discrimination responses per participant (See Table 7 for AXB contrasts). Of the five consecutive repetitions produced by our speakers for each sequence (see section 2.1.1), only the third token was used in Experiment 3. To avoid triad sequence bias, tokens were organised into a Latin square. Each contrast test contained tokens from each of the three speakers organised into six speaker sequences (123, 132, 213, 231, 312, 321) and each of the speaker sequences was organised into four token sequences (AAB, ABB, BAA, BBA). AXB

discrimination tests were drawn at random from this pool of 24 individual triads. The AXB discrimination task was presented using Psyscope X (Bonatti, 2010) and discrimination responses and response times were both recorded. Response time was measured after AXB triads were presented completely. Response times were recorded because studies have shown that response time is a better indicator of phonotactic probability than response accuracy (Edwards et al., 2004). The AXB discrimination triads were spaced with a 1500 millisecond inter-trial interval. Within AXB trials, the interstimulus interval (ISI) was 1000 milliseconds, to ensure phonological rather than phonetic processing (Werker & Tees, 1984). The order in which the five blocks were presented was counterbalanced across participants, reducing the possible influence of test-fatigue or task-specific learning systematically affecting one or more of the individual tasks. If participants failed to respond to trials within a 2000 millisecond response window, they were presented with an on-screen message in Japanese, politely asking them to respond faster. Missed trials were replayed at a random time during the remainder of the experiment.

3.2 Predictions

In Experiment 2: Discrimination, we test the following hypotheses based on our extension of PAM-L2:

H4) Participants will be more accurate at discriminating between TC contrasts (/iʃi/-/iti/ and /uʃu/-/usu/) than CG contrasts (/itʃi/-/iti/, /iʃi/-/isi/ and /utsu/-/utu/). Listeners will be better at discriminating between strings that can be mapped to two native categories (TC) than those strings that present as instances of a single category (CG) but vary in how listeners perceive them to 'fit' to that category.

H5) Pairs of phonotactically legal and illegal strings will vary in their discriminability. This variance will depend on the perceptual difference between illegal and legal strings. We propose that the effect size calculations from the *t*-tests conducted on the mean goodness-of-fit ratings in Experiment 1 (Table 6) provide an indication of the perceptual distance of the English phoneme strings, because they standardise differences between the means of the goodness-of-fit ratings assigned to phonotactically legal and phonotactically illegal strings.

H6) Participants will take longer to respond to CG contrasts than TC contrasts. We hypothesise that this is due to CG discrimination trials constituting a heavier cognitive load.

3.3 Discrimination: Results

A series of one sample *t*-tests was run to measure discrimination accuracy against chance (50%). All *t*-tests returned a significant result, as shown in Table 9, indicating that the discrimination performance for all contrasts differed from chance performance.

As predicted by our hypothesis (H4), the participants were more accurate in discriminating TC contrasts than they were discriminating between CG contrasts. Indeed, the native Japanese participants displayed excellent discrimination accuracy for both the two TC contrasts /iʃi/-/iti/ (97% correct discrimination) and /uʃu/-/usu/ (90% correct discrimination), while the discrimination accuracy for the three CG legal-illegal contrasts, /itʃi/-/iti/ (80% correct discrimination) /iʃi/-/isi/ (75% correct discrimination), and /utsu/-/utu/, (68% correct discrimination), was poorer.

A one-way, between groups ANOVA confirmed that there was a significant difference between the contrasts, $F(4) = 19.279$, $p = < 0.001$, with a large effect size (Cohen, 1988), partial $\eta^2 = 0.473$. Further post-hoc Bonferroni-corrected analyses confirmed that participants were more accurate in discriminating between predicted TC contrasts than predicted CG contrasts (see Table 10 for the complete results from the Bonferroni-corrected multiple comparisons). Indeed, there are significant differences between the discriminability of the /iʃi/-/iti/ contrast and all CG contrasts (pairs of legal and illegal strings): /itʃi/-/iti/, $p = 0.019$, /iʃi/-/isi/, $p = 0.001$, and /utsu/-/utu/, $p = < 0.001$. These results support H4 because they reveal a significant difference in discriminability between the /iʃi/-/iti/ TC contrast, and all CG contrasts. The /uʃu/-/usu/ contrast, on the other hand, only differed significantly from /iʃi/-/isi/, $p = 0.047$, and /utsu/-/utu/, $p = 0.001$, and not from /itʃi/-/iti/, $p = 0.55$. This is possibly the result of the large variability in the /itʃi/-/iti/ results, as can be seen in the standard deviation shown in Figure 2.

To test the hypothesis that participants would take longer to respond to CG tests than TC tests (H6), we examined the response times of participants across the five AXB discrimination tasks using a between groups ANOVA. The results indicate that there is a significant difference between the response times of AXB tests with a medium effect size, $F(4) = 39.779$, $p = < 0.001$, partial $\eta^2 = 0.05$. As a comparison, a between groups ANOVA was calculated to determine the effect of different participants on response time, this ANOVA exposed a much larger effect size, $F(9) = 128.870$, $p = < 0.001$, partial $\eta^2 = 0.279$.

To further test H6, a Bonferroni analysis was calculated to determine response time pairwise comparisons between AXB tests. The Bonferroni post-hoc comparisons revealed that participants took significantly longer to respond to CG discrimination tests, (/itʃi/-/iti/: *M* = 1130, *SD* = 161); /iʃi/-/isi/: *M* = 1207, *SD* = 151; and /utsu/-/utu/: *M* = 1193, *SD* = 174) than to TC discrimination tests (/iʃi/-/iti/: *M* = 1022, *SD* = 130; /uʃu/-/usu/: *M* = 1071, *SD* = 147).

The Bonferroni-corrected post-hoc comparisons showed that there was no significant difference between the response times for those trials that involved TC discrimination (/iʃi/-/iti/ and /uʃu/-/usu/). Similarly, there was no significant difference between /iʃi/-/isi/ and /utsu/-/utu/ CG discrimination trials. However, there was a significant difference between the response times of /itʃi/-/iti/ and other CG discrimination trials, /iʃi/-/isi/ and /utsu/-/utu/. Together, these results reveal that participants required significantly less time to respond to the /itʃi/-/iti/ CG discrimination task compared to the other two CG tasks /iʃi/-/isi/ and /utsu/-/utu/. These results, as well as the discrimination accuracy results, show that the /itʃi/-/iti/ contrast is perceptually more distinct than /iʃi/-/isi/ and /utsu/-/utu/.

3.3.1 Correlations between Discrimination Accuracy and Response Time

A Spearman's correlation was computed to assess the relationship between average time taken to respond to AXB trials and correct answers. There was a significant negative correlation between average response time and AXB trial scores (r = -.617, p = < 0.001) with a large effect size (d = -8.611). The scatter plot in Figure 4 illustrates this negative relationship. This is an important finding because it confirms the usefulness of measures of response time as an indicator of perceptual distance and, in turn, task difficulty. Indeed, contrasts that are more easily discriminable achieve higher accuracy and lower response times while the opposite is true of contrasts that are more difficult to discern. This also allows for a more fine-grained analysis than task accuracy because each participant only provides 5 data points for accuracy (1 per contrast), whereas each participant returns 300 response times (60 per contrast).

3.4 Experiment 2: Discussion

Experiment 2 tested the discrimination accuracy and response time of Japanese participants in discriminating TC and CG string contrasts. The discrimination results are consistent with our H4 and show that the participants were significantly less accurate in discriminating between CG contrasts than TC contrasts, such that the discrimination accuracy of TC /iʃi/-/iti/ contrast was much greater than the accuracy of the three CG tests. Interestingly,

while the TC /uʃu/-/usu/ contrast was more accurately discriminated than the /iʃi/-/isi/ and /utsu/-/utu/ contrasts, discrimination accuracy for this contrast did not differ significantly from that of the CG /itʃi/-/iti/ contrast. This lack of a difference in discrimination accuracy between TC /uʃu/-/usu/ and CG /itʃi/-/iti/ contrasts is likely the result of a large variation in the latter. However, despite the lack of a significant difference in discrimination accuracy between /uʃu/-/usu/ and /itʃi/-/iti/, the results generally support the prediction that the participants would more accurately discriminate between TC contrast pairs than between CG contrast pairs. This is consistent with our predictions that phonotactically illegal /CV/ combinations will be perceived as similar to their perceptually nearest phonotactically legal Japanese string, increasing the likelihood that they will be perceived inaccurately by native listeners. Additionally, we find support for our predictions in the fact that the Bonferroni comparisons found no statistical difference in discrimination accuracy within the subsets of TC versus CG contrasts.

The Bonferroni correction conducted on the between groups ANOVA on AXB trial discrimination accuracy and response times also support our hypotheses. While, the significant relationship, with a large effect size (see Table 11), shows that, of the three AXB trials consisting of pairs of phonotactically legal and illegal /VCV/ strings indicates that the /itʃi/-/iti/ trial responses were significantly faster than those responses to the /iʃi/-/isi/ and /utsu/-/utu/ trials, the results also show that the /iʃi/-/isi/ and /utsu/-/utu/ contrasts do not differ. We propose that the observed differences in the response times is due to differences in the cognitive loads of discrimination between different string types (TC versus CG). Indeed, given that pairwise comparisons between /iʃi/-/iti/ and /uʃu/-/usu/, and /iʃi/-/isi/ and /utsu/-/utu/ failed to achieve significance, this suggests that the response times classify tests into the following three categories:

1) Pairs of English strings that are mapped to two phonotactically legal contrasts in Conservative Japanese (/iʃi/-/iti/ and /uʃu/-/usu/).

2) Pairs of English strings that violate Japanese phonotactics but frequently occur as contrasts in Innovative Japanese (/itʃi/-/iti/).

3) Pairs of English strings that violate Japanese phonotactics and infrequently occur as contrasts in Innovative Japanese (/iʃi/-/isi/ and /utsu/-/utu/).

4. General Discussion

The present paper examines the influence of Japanese phonotactics on the perception of English consonant strings by native speakers of Japanese with very limited L2-English proficiency. To our knowledge, this is the first study to examine the influence of native (Japanese) consonant-vowel co-occurrence restrictions on the perception of non-native (English) consonants. The results provide clear evidence that the attunement of perceptual systems to native phonotactic restrictions results in a decrease of discriminability of consonants in environments that violate native phonotactics. The present paper firstly extends the Perceptual Assimilation Model (Best, 1995; Best & Tyler, 2007) to include phonotactics in order to account for the crucial role of phonotactics in cross-language and L2 segmental speech perception. According to this novel framework, strings of native segments in sequences that violate native phonotactics follow similar patterns of assimilation as individual, non-native segments, as outlined in models like PAM and PAM-L2. The present paper then tests these predictions in two experiments focusing on the perception of English /VCV/ strings by near-monolingual Japanese listeners. Experiment 1 (Section 2), showed that Japanese listeners categorised English /VCV/ strings in a manner consistent with our predictions, assimilating phonotactically illegal strings to the perceptually nearest phonotactically legal /VCV/ string. Experiment 2 (Section 3) showed that Japanese listeners are more accurate and have faster response times when discriminating between two phonotactically legal strings than between legal and illegal strings that are assimilated into the same native (phonotactically legal) string.

These results show the necessity for an extension of PAM-L2 as they indicate that the phonological status of a non-native segment is not the only important feature in predicting non-native categorisation patterns and discrimination accuracy. Indeed, the native Japanese participants' discrimination accuracy differs even between /uʃu/-/usu/ and /iʃi/-/isi/ AXB tests, where the same pairs of English phones /s/ and /ʃ/ have different levels of discrimination accuracy depending on the phonotactics of the string in which they are presented. PAM-L2 (as opposed to the current extension) would predict that Japanese listeners would discriminate between /s/ and /ʃ/ in varying contexts with equal accuracy, however, as indicated by the different results from the /uʃu/-/usu/ and /iʃi/-/isi/ AXB discrimination tasks, phonotactic context does indeed influence discriminability. In contrast to PAM-L2, our extension predicts that while English /su/ and /ʃu/ are mapped to existing native Japanese string categories (す /su/ and しゅ /ʃu/), English /si/ by contrast is perceived as a less canonical version of its perceptually

nearest legal Japanese sequence (ㄴ /ʃi/). Japanese co-occurrence restrictions thus provide new insights into the influence of phonotactics on non-native language perception. This suggests that listeners' perceptual systems are focused on discriminating between those sequences of sounds that occur frequently in L1 discourse.

We suggest that these results indicate that weakening phonotactic restrictions resulting from loanword influence have exposed even monolingual Japanese speakers to /CV/ strings that were once absent from the Japanese language, allowing for a process of perceptual tuning whereby an individual's linguistic perceptual systems adapt to focus on novel phonologically and phonetically meaningful information. Indeed, the results provide support for the observation that contrasting Japanese /CV/ sequences that occur at a high frequency have undergone perceptual broadening (for instance, /ti/) and that this process has enhanced the perceptual distance between commonly occurring strings, making them easier to distinguish between.

This effect of weakening phonotactics inducing perceptual reattunement is supported by the fact that participants more accurately discriminated between the /itʃi/-/iti/ contrast than /iʃi/-/isi/ and /utsu/-/utu/ contrasts. Indeed, this appears consistent with our prediction that listeners are better tuned to discriminate /iti/ due to the probable increased frequency at which /ti/ occurs in Japanese discourse based on Tamaoka and Makioka's (2004) corpus analysis and Itô & Mester's (1999) core periphery model. Differences of occurrence frequency between /VCV/ sequences thus account for a difference in the perceptual tuning of sequences undergoing phonotactic weakening, creating perceptual distance between frequently occurring /CV/ sequences, and perceptual nearness between infrequently occurring sequences and their closest native category (see Bundgaard-Nielsen & Baker, 2014 for evidence that this also occurs on a segmental level). This perceptual tuning is reflected in the TC discrimination results, the discrimination response times and in the effect size calculations on *t*-tests, from Experiment 1, which measure between-category Goodness-of-fit ratings (see Table 8). In all of these results, the /itʃi/ and /iti/ pair is intermediate between TC and CG pairs.

Consistent with research on phonotactic probability (Edwards et al., 2004) and perception of accented speech (Munro & Derwing, 1995), the results of the present paper indicate that response time is a more accurate indicator of perceptual distance than accuracy in studies with highly consistent L2 phone or string discrimination. Indeed, we propose that response time is directly influenced by the perceptual distance between phonotactically legal

and illegal non-native phoneme strings, and that perceptually more similar contrasts require more time to process than perceptually distant contrasts. Importantly however, in the experiments presented here, discrimination test response times were not only significantly different between TC and CG tests but also between /itʃi/-/iti/ and other CG tests (Table 11), indicating that, for near-monolingual Japanese listeners, the perceptual distance between /ti/-/tʃi/ is greater than that of /si/-/ʃi/ and /tu/-/tsu/, despite similar categorisation patterns in Experiment 1. We maintain that the perceptual distance between /ti/-/tʃi/ is greater than that of /si/-/ʃi/ and /tu/-/tsu/ due to the different rates of phonologisation of these allophones which is in turn likely the result of weakening phonotactic constraints over time.

The results from the present study are also consistent with the notion that extended response times are the manifestation of an increased cognitive load due to the difficulty experienced in differentiating between perceptually similar phones. (e.g., Adank et al., 2009; Munro & Derwing, 1995; Schmid & Yeni-Komshian, 1999). While the exact nature of an increased cognitive load can only be speculated upon, it may involve either Type 1 or Type 2 processing (Stanovich & West, 2000). Type 1 processing involves more autonomous, unconscious processes, whereas Type 2 processing is reliant on top down cognition and requires effortful and conscious control by the central executive. A Type 2 explanation might involve listeners replaying the stimuli in the phonological loop for further analysis or a similar process within those systems associated with working memory. Alternatively, non-native perception might require an extra layer of more autonomous, Type 1 processing which may account for the extra time required to assimilate non-native phones. For example, when Japanese listeners are exposed to /isi/, they autonomously perceive three separate phones but their perceptual system recognises that this string is a violation of native phonotactics and perceptually repairs the string to adhere to the nearest phonotactically legal string. The recognition and repair sequence in this example might account for the response time differences observed in the present paper.

Our extension of PAM-L2 (Best & Tyler, 2007) recommends the consideration of larger chunks of signal in perceptual assimilation models. However, we do not maintain that larger chunks of signal constitute basic coding units of speech. Our amendment must then consider an alternative explanation for the influence of phonotactics on perception. We submit that when listeners are exposed to illicit sequences of speech, their perceptual systems sometimes make erroneous predictions as to the nature of the incoming signal. These predictions are based on

the listener's linguistic experience where they try to match unfamiliar speech to that which is both perceptually similar and maintains a high transitional probability in their L1. The likelihood that a listener will make such an erroneous prediction is therefore based upon the perceptual distance between the incoming signal and its nearest match as well as the frequency at which the listener has been exposed to the illicit sequence. The proposal that the frequency with which a phonological sequence occurs in a given language motivates perceptual tuning has been established in a range of studies designed to measure implicit phonotactic knowledge. Studies examining the relationship between phonotactic probability and perception have found that high frequency sequences are recognised with greater speed and accuracy than those that occur at low frequencies (e.g., Vitevitch & Luce, 1998; Vitevitch et al., 1999). Studies have demonstrated that listeners are perceptually biased to recognise ambiguous phonemes in phonological sequences to match high frequency sequences (Pitt & McQueen, 1998). Phonotactic probability has also been found to influence non-native speech production. For instance, researchers have found that adult speakers are faster and more accurate at repeating nonce words that contain /CV/ and /VC/ sequences that occur frequently in their native languages than those that do not (Vitevitch & Luce, 1999; Vitevitch et al., 1997). However, the relationship between repetition accuracy and phonotactic probability in these studies is not as robust nor as consistently replicated across experiments as is the relationship between response time and phonotactic probability (Edwards et al., 2004).

While the results of this paper clearly show that Japanese co-occurrence restrictions influence L2 perception, there are a number of limitations of this study that should be addressed. Firstly, all of our participants have not only been exposed to a considerable amount of the English language but have undergone English language instruction. Ideally, our participants would have had little English exposure, however, Japan has a policy of compulsory English instruction so this was unavoidable with adult participants. Additionally, our extension of PAM-L2 proposes that the frequency in which sounds co-occur in a language influences their perception, yet four of our five legal strings constitute existing lexical items in Japanese and numerous studies have shown an effect of frequency and semantics on speech perception. As mentioned previously it is difficult to randomly combine mora without creating existing Japanese words, this was unavoidable given the co-occurrence restrictions we were aiming to test. However, of the four Japanese words in our experimental strings, /itʃi/ 'one' is likely the most frequently occurring and if this was influencing our results then we would expect that /iti/

and /itʃi/ would be more difficult to discriminate between than other GC pairs which was not the case.

In conclusion, the present paper demonstrates that Japanese phonotactics influence perception of English consonants. Indeed, Japanese co-occurrence restrictions have a direct and systematic influence on the ability of listeners to differentiate between English strings, consistent with our predictions, and we contend that L2 segmental perception may be shaped not just by the L1 segmental inventory, but also by L1 phonotactics, and finally that perceptual tuning to non-native di-phones depends on relative exposure frequencies in the input. We argue that studies such as that of the present paper have clear theoretical implications given that most contemporary models of non-native and cross-language speech perception, like PAM-L2, predict and account for the varying degrees of success that learners have with non-native phonetic and phonological contrasts on a segmental level alone but offer no framework or predictions for the ways in which native phonotactics may play a role in non-native segmental perception.

Acknowledgements

We thank our participants as well as Eri Suzuki and Akiko Murata who helped to recruit and organise participants. We also thank Rosey Billington, Sally Bowman, Katie Jepson, and Eleanor Lewis for lending their voices to record tokens. Finally, we would like to thank three anonymous reviewers and the editors at Applied Psycholinguistics whose invaluable suggestions resulted in significant improvements to the manuscript.

References

Adank, P., Evans, B. G., Stuart-Smith, J., & Scott, S. K. (2009). Comprehension of familiar and unfamiliar native accents under adverse listening conditions. *Journal of Experimental Psychology: Human Perception and Performance, 35*(2), 520.

Best, C. T. (1994). The emergence of native-language phonological influences in infants: A perceptual assimilation model. In J. C. Goodman & H. C. Nusbaum (Eds.), *The development of speech perception: The transition from speech sounds to spoken words* (pp. 167-224). Cambridge, MA: MIT Press.

Best, C. T. (1995). A direct-realist view of cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in crosslanguage research* (pp. 171-204). Timonium: York Press.

Best, C. T., & Strange, W. (1992). Effects of language-specific phonological and phonetic factors on cross-language perception of approximants. *Journal of Phonetics, 20,* 305-330.

Best, C. T., & Tyler, M. D., (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In J. Munro & O. S. Bohn (Eds.), *Second language speech learning: The role of language experience in speech perception and production* (pp. 13-34). Amsterdam: John Benjamins.

Blevins, J. (1995). The Syllable in Phonological Theory. In J. Goldsmith (Ed.), *Handbook of phonological theory* (pp. 206-44). London: Basil Blackwell.

Bloch, B. (1950). Studies in Colloquial Japanese IV Phonemics. *Language, 26*(1), 86-125.

Boersma, P., & Weenink, D. (2015). Praat (Version 5.4. 08) [Computer software]. Retrieved March 12, 2015, from http://www.praat.org/.

Bonatti, L. L. (2010). Psyscope X (Build 77) [computer software]. Retrieved January 15, 2015, from http://psy.ck.sissa.it/.

Bundgaard-Nielsen R. L., & Baker B. (2014). Frequency in the input affects perception of phonological contrasts for native speakers. In J. Hay & E. Parnell (Eds.), *Proceedings of the 15th Australasian International Speech Science and Technology Conference* (pp. 205-208). Christchurch, New Zealand. University of Canterbury: Australasian Speech Science and Technology Association.

Bundgaard-Nielsen, R. L., Best, C. T., & Tyler, M. D. (2011). Vocabulary size is associated with second-language vowel perception performance in adult learners. *Studies in Second Language Acquisition*, *33*(3), 433-461.

Cabinet of Japan's Ministry of Education, Culture, Sports, Science and Technology (1991). 外来語の表記. Retrieved from http://www.mext.go.jp/b_menu/hakusho/nc/k19910628002/k19910628002.html.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hill*SD*ale: Erlbaum.

Cuetos, F., Hallé, P., Dominguez, A., & Segui, J. (2011). Perception of prothetic /e/ in #sC utterances: *Gating data. Oral communication at the 17th ICPhS, 17*, 17-21.

Davidson, L., & Shaw, J. A. (2012). Sources of illusion in consonant cluster perception. *Journal of Phonetics, 40*(2), 234-248.

Dupoux, E., Kakehi, K., Hirose, Y., Pallier, C., & Mehler, J. (1999). Epenthetic vowels in Japanese: A perceptual illusion?. *Journal of experimental psychology: human perception and performance, 25*(6), 1568.

Dupoux, E., Parlato, E., Frota, S., Hirose, Y., & Peperkamp, S. (2011). *Where do illusory vowels come from?. Journal of Memory and Language, 64*(3), 199-210.

Edwards, J., Beckman, M. E., & Munson, B. (2004). The interaction between vocabulary size and phonotactic probability effects on children's production accuracy and fluency in nonword repetition. Journal of Speech, *Language, and Hearing Research, 47*(2), 421-436.

Flege, J. E. (1995). Second language speech learning – theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 233-277). Baltimore: York Press.

Hallé, P. A., & Best, C. T. (2007). Dental-to-velar perceptual assimilation: A cross-linguistic study of the perception of dental stop + /l/ clusters. *The Journal of the Acoustical Society of America, 121*(5), 2899-2914.

Hallé, P. A., Dominguez, A., Cuetos, F., & Segui, J. (2008). Phonological mediation in visual masked priming: Evidence from phonotactic repair. *Journal of Experimental Psychology: Human Perception and Performance, 34*(1), 177.

Hallé, P. A., Segui, J., Frauenfelder, U., & Meunier, C. (1998). Processing of illegal consonant clusters: A case of perceptual assimilation?. *Journal of experimental psychology: Human perception and performance*, *24*(2), 592.Itô, J., & Mester, A. (1995). The core-periphery structure of the lexicon and constraints on reranking. *Papers in optimality theory*, *18*, 181-209.

Itô, J., & Mester, A. (1999). The Phonological Lexicon. In N. Tsujimura (Ed.), *The Handbook of Japanese Linguistics* (pp. 62-100). Oxford: Blackwell Publishers.

Kabak, B., & Idsardi, W. J. (2007). Perceptual distortions in the adaptation of English consonant clusters: Syllable structure or consonantal contact constraints?. *Language and Speech*, *50*(1), 23-52.

Kahn, D. (1976). *Syllable-based generalizations in English phonology* (Doctoral dissertation, Massachusetts Institute of Technology).

Kawahara, S., Ono, H., & Sudo, K. (2006). Consonant co-occurrence restrictions in Yamato Japanese. *Japanese/Korean Linguistics*, *14*, 27-38.

Kuhl, P. K. (1993). Innate predispositions and the effects of experience in speech perception: The native language magnet theory. In B. de Boysson-Bardies, S. de Schonen, P. Jusczyk, P. MacNeilage & J. Morton (Eds.), *Developmental neurocognition: Speech and face processing in the first year of life* (pp. 259-274). Hingham: Kluewr Academic Press.

Kuhl, P. K., Conboy, B. T., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M., & Nelson, T. (2008). Phonetic learning as a pathway to language: new data and native language magnet theory expanded (NLM-e). *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1493), 979-1000.

Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., & Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science*, *255*(5044), 606-608.

Lewkowicz, D. J. (2014). Early experience and multisensory perceptual narrowing. *Developmental psychobiology*, *56*(2), 292-315.

Munro, M. J., & Derwing, T. M. (1995). Processing time, accent, and comprehensibility in the perception of foreign-accented speech. *Language and Speech*, *38*, 289 –306.

Mutsukawa, M. (2009). *Japanese Loanword Phonology: The Nature of Inputs and the Loanword Sublexicon*. Tokyo: Hituzi Syobo.

Pintér, G. (2015). The emergence of new consonant contrasts. In H. Kubozono (Ed.), *Handbook of Japanese Phonetics and Phonology* (pp. 121-165). Berlin: De Gruyter Mouton.

Pitt, M. A., & McQueen, J. M. (1998). Is compensation for coarticulation mediated by the lexicon?. *Journal of Memory and Language*, *39*(3), 347-370.

Schmid, P. M., & Yeni-Komshian, G. H. (1999). The effects of speaker accent and target predictability on perception of mispronunciations. *Journal of Speech, Language, and Hearing Research*, *42*(1), 56-64.

Scott, L. S., & Monesson, A. (2010). Experience-dependent neural specialization during infancy. *Neuropsychologia*, *48*(6), 1857-1861.

Scott, L. S., Pascalis, O., & Nelson, C. A. (2007). A domain-general theory of the development of perceptual discrimination. *Current Directions in Psychological Science*, *16*(4), 197-201.

Stanovich, K. E., & West, R. F. (2000). Advancing the rationality debate. *Behavioral and brain sciences*, *23*(5), 701-717.

Tamoka, K., & Makioka, S. (2004). Frequency of occurrence for units of phonemes, morae, and syllables appearing in a lexical corpus of a Japanese newspaper. Behaviour Research Methods, Instruments & Computers, 36(3), 531-547.

Tsujimura, N. (2013). *An Introduction to Japanese Linguistics* (3rd ed.). Malden: Blackwell Publishers Inc.

Vance, T. J. (1987). *An Introduction to Japanese Phonology*. Albany: State University of New York Press.

Vitevitch, M. S., & Luce, P. A. (1998). When words compete: Levels of processing in perception of spoken words. *Psychological science*, *9*(4), 325-329.

Vitevitch, M. S., Luce, P. A., Charles-Luce, J., & Kemmerer, D. (1997). Phonotactics and syllable stress: Implications for the processing of spoken nonsense words. *Language and speech*, *40*(1), 47-62.

Vitevitch, M. S., Luce, P. A., Pisoni, D. B., & Auer, E. T. (1999). Phonotactics, neighborhood activation, and lexical access for spoken words. *Brain and language*, *68*(1), 306-311.

Wagner, M., Shafer, V. L., Martin, B., & Steinschneider, M. (2012). The phonotactic influence on the perception of a consonant cluster /pt/ by native English and native Polish listeners: a behavioral and event related potential (ERP) study. *Brain and language*, *123*(1), 30-41.

Werker, J. F., Gilbert, J. H., Humphrey, K., & Tees, R. C. (1981). Developmental aspects of cross-language speech perception. *Child development*, *52*(1), 349-355.

Werker, J. F., Pons, F., Dietrich, C., Kajikawa, S., Fais, L., & Amano, S. (2007). Infant-directed speech supports phonetic category learning in English and Japanese. *Cognition*, *103*(1), 147-162.

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant behavior and development*, *7*(1), 49-63.

Figures

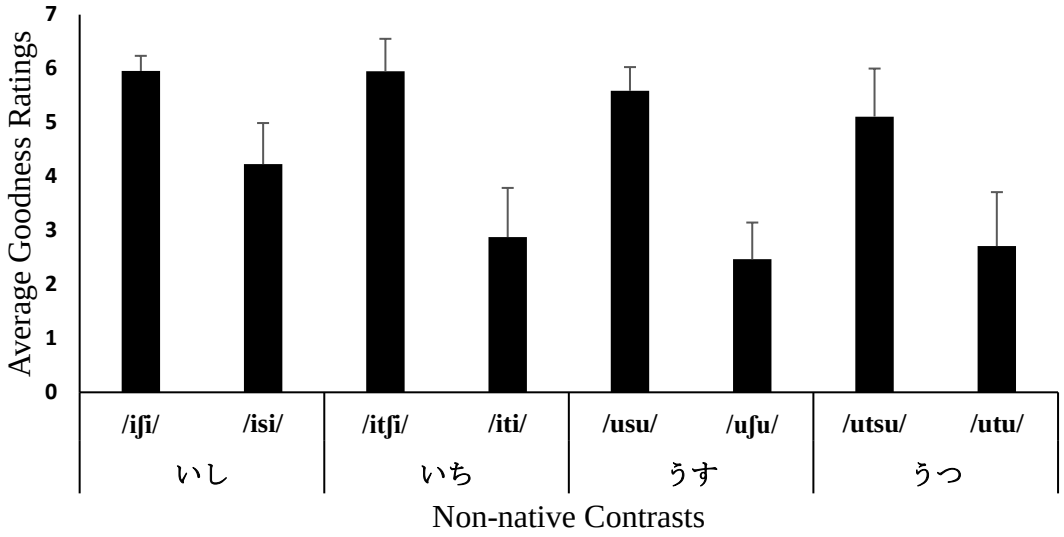

Figure 1. Average goodness-of-fit results. Error bars indicate standard deviation.

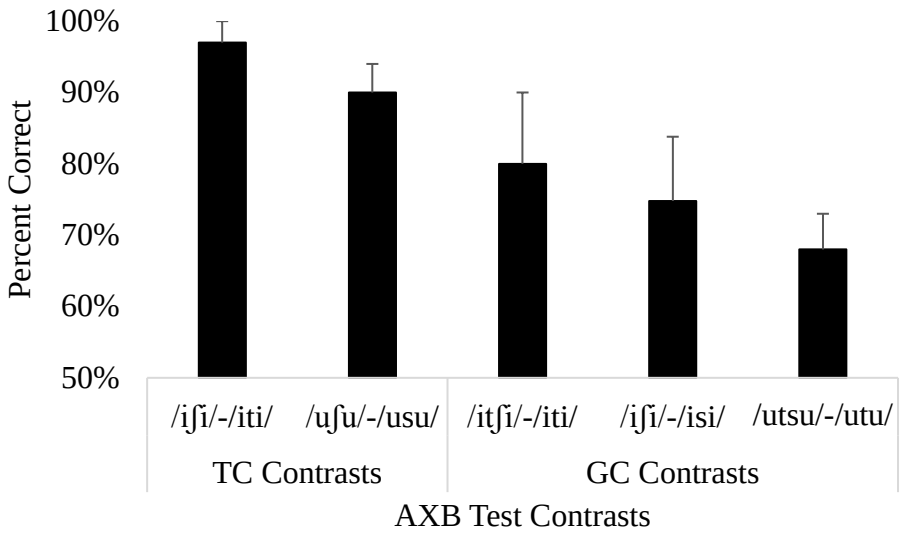

Figure 2. Percent correct discrimination for the five contrasts. Error bars represent standard deviation.

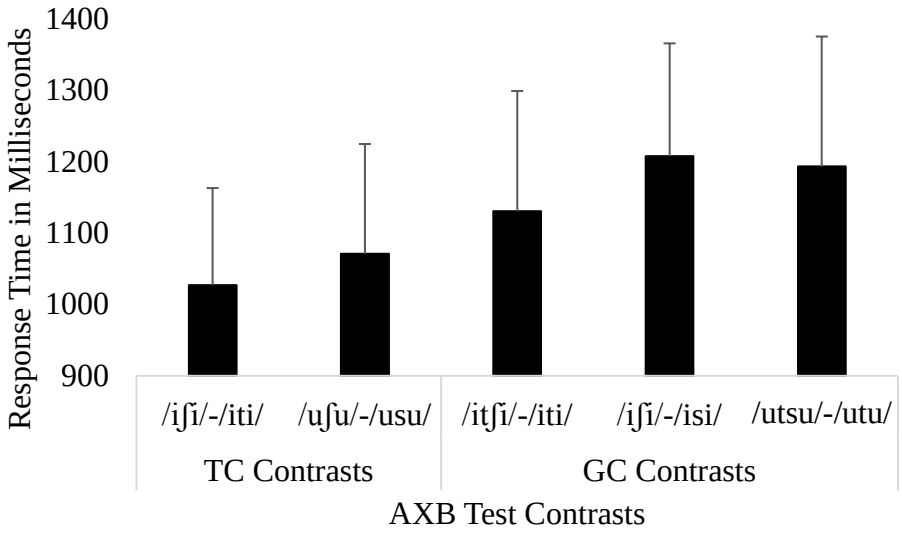

Figure 3. Average response time for each AXB test. Error bars represent standard deviation.

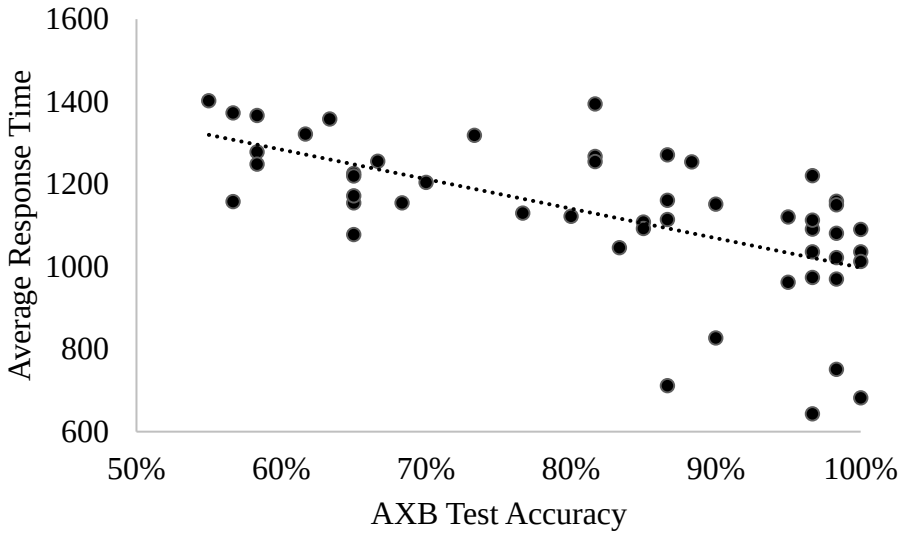

Figure 4. Scatter plot showing the negatively correlated relationship between average response time and AXB test accuracy.

Tables

Table 1. AXB test contrasts used in Experiment 2. Illegal strings are represented by bold type.

/iʃi/	-	/iti/
/uʃu/	-	/usu/
/itʃi/	-	**/iti/**
/iʃi/	-	**/isi/**
/utsu/	-	**/utu/**

Table 2. Selection of possible CV combinations in Conservative and Innovative Japanese. Adapted from Bloch (1950, p. 123). Bold type represents innovative combinations.

	/s/	**/ʃ/**	**/t/**	**/tʃ/**	**/ts/**
/a/	sa	ʃa	ta	tʃa	**tsa**
/i/	**si**	ʃi	**ti**	tʃi	**tsi**
/u/	su	ʃu	**tu**	tʃu	tsu
/e/	se	**ʃe**	te	tʃe	**tse**
/o/	so	ʃo	to	tʃo	**tso**

Table 3. Katakana graphemes used for transcribing CV combinations. Innovative combinations—represented by bold type—are standardised forms proposed by the Cabinet of Japan's Ministry of Education, Culture, Sports, Science and Technology (1991). Here, /si/ is left blank because while スィ is offered as a suggestion, the report indicates that the transcription of /si/ is up to the discretion of the individual.

	/s/	**/ʃ/**	**/t/**	**/tʃ/**	**/ts/**
/a/	サ	シャ	タ	チャ	**ツァ**
/i/		シ	**ティ**	チ	**ツィ**
/u/	ス	シュ	**トゥ**	チュ	ツ
/e/	セ	**シェ**	テ	チェ	**ツェ**
/o/	ソ	ショ	ト	チョ	**ツォ**

Table 4. Categories presented in hiragana and romaaji.

Hiragana	いし	いち	うす	うつ
Romanisation	ishi	ichi	usu	utsu

Table 5. Categorisation patterns. Initial figures represent the percentage of responses to a given category and figures in parentheses present the average goodness-of-fit ratings on a scale from 1 (bad) to 7 (very good).

		Categories			
		いし **ishi**	いち **ichi**	うす **usu**	うつ **utsu**
Stimuli	**/iʃi/**	100% (5.96)			
	/isi/	100% (4.23)			
	/itʃi/		100% (5.95)		
	/iti/		100% (2.88)		
	/usu/			98% (5.59)	2%
	/uʃu/			98% (2.47)	2%
	/utsu/			1%	99% (5.11)
	/utu/			1%	99% (2.71)

Table 6. Within-category paired sample *t*-test results.

Category	Paired Samples	Mean Diff.	Standard Deviation	*t*	*p*	Cohen's *d*
いし	/iʃi/-/isi/	1.726*	1.028	6.883	< 0.001	3.017
いち	/itʃi/-/iti/	3.074*	0.792	9.454	< 0.001	3.994
うす	/uʃu/-/usu/	3.121*	0.815	12.111	< 0.001	5.440
うつ	/utsu/-/utu/	2.401*	1.231	6.169	< 0.001	2.548

Table 7. AXB discrimination trials organised into PAM-L2 based categories

Two Category Difference	Category Goodness Difference
/iʃi/ & /iti/	/itʃi/ & /iti/
/uʃu/ & /usu/	/iʃi/ & /isi/
	/utsu/ & /utu/

Table 8. Goodness-of-fit ratings for prototypical and non-prototypical strings.

Category	Prototypical	Score	Non-Prototypical	Score
いし	/iʃi/	5.96	/isi/	4.23
いち	/itʃi/	5.95	/iti/	2.88
うす	/usu/	5.56	/uʃu/	2.47
うつ	/utsu/	5.11	/utu/	2.71

Table 9. One sample *t*-tests measuring dissimilarity between AXB discrimination test results and chance results (50%).

AXB test	Accuracy	*t*	*p*
/iʃi/-/iti/	97%	29.986	< 0.001
/uʃu/-/usu/	90%	18.864	< 0.001
/itʃi/-/iti/	80%	5.494	< 0.001
/iʃi/-/isi/	75%	5.362	< 0.001
/utsu/-/utu/	68%	6.306	< 0.001

Table 10. Bonferroni pairwise comparisons of AXB test results.

	Pairwise Comp.	Mean Diff.	S.E.	*p*	95% Confidence Interval	
					Lower Bound	Upper Bound
/iʃi/-/iti/	/uʃu/-/usu/	4.1	3.096	1	-5.04	13.24
	/itʃi/-/iti/	10.2*	3.096	0.019	1.06	19.34
	/iʃi/-/isi/	13.3*	3.096	0.001	4.16	22.44
	/utsu/-/utu/	17.3*	3.096	< 0.001	8.16	26.44
/uʃu/-/usu/	/itʃi/-/iti/	6.1	3.096	0.55	-3.04	19.34
	/iʃi/-/isi/	9.2*	3.096	0.047	0.06	18.34
	/utsu/-/utu/	13.2*	3.096	0.001	4.06	22.34
/itʃi/-/iti/	/iʃi/-/isi/	3.1	3.096	1	-12.24	6.04
	/utsu/-/utu/	7.1	3.096	0.265	-2.04	16.24
/iʃi/-/isi/	/utsu/-/utu/	4	3.096	1	-5.14	13.14

Table 11. Bonferroni pairwise comparisons of AXB test response time results.

					95% Confidence Interval	
	Pairwise Comparison	**Mean Diff.**	**S.E.**	***p***	**Lower Bound**	**Upper Bound**
/iʃi/-/iti/	/uʃu/-/usu/	23.31	15.997	1	-68.19	21.57
	/itʃi/-/iti/	82.91*	15.997	< 0.001	-127.80	-38.03
	/iʃi/-/isi/	160.06*	15.997	< 0.001	-204.95	-115.18
	/utsu/-/utu/	145.6*	15.997	< 0.001	-190.48	-100.72
/uʃu/-/usu/	/itʃi/-/iti/	59.61 *	15.997	0.002	-104.49	-14.73
	/iʃi/-/isi/	136.74*	15.997	< 0.001	-181.64	-91.88
	/utsu/-/utu/	122.3*	15.997	< 0.001	-167.18	-77.42
/itʃi/-/iti/	/iʃi/-/isi/	77.15*	15.997	< 0.001	-122.03	-32.27
	/utsu/-/utu/	62.69*	15.997	0.001	-107.57	-17.81
/iʃi/-/isi/	/utsu/-/utu/	14.46	15.997	1	-30.42	59.34