ISSN 0286-1291. Doshisha Studies in English, No. 96. March, 2016. pp. 61-111

Errata
p. 66, line 3 & line 11, Cooper et al. → Dupoux et al.
p. 79, Table 2. (caption) The percentage of nouns with antepenultimate accent among all accented nouns→The percentage of nouns with antepenultimate accent among trimoraic accented nouns

Is Japanese listeners' perception of English stress influenced by the antepenultimate accent in Japanese? Comparison with English and Korean listeners

Mariko Sugahara

Abstract

Sugahara (2011) has demonstrated that Japanese listeners were strongly biased towards iambic rhythm when judging the stress pattern of English disyllabic words presented as auditory stimuli where F0 was not a reliable cue to distinguish trochaic forms, e.g., *TRANSplant*, and iambic forms, e.g., *transPLANT*. Both the disyllabic nouns and the verbs used in Sugahara (2011) have their antepenultimate mora, i.e., their final syllable nucleus vowel, accented once they are adopted into Japanese as loanwords, and the hypothesis tested in this study is that Japanese listeners' iambic bias is due to their English stress perception being influenced by the antepenultimate accent in Japanese. A forced identification experiment was carried out to compare Japanese listeners' perception of English stress with native English listeners' and Seoul Korean listeners' perception using auditory stimuli similar to those used in Sugahara (2011). Unlike Japanese, the English vocabulary is dominated by word-initial primary stress, and there is no lexical stress/accent contrast in Seoul Korean. When the F0 information of the stimuli was synthesized in such a way as not to disambiguate the trochaic and iambic patterns, Japanese listeners showed a strong iambic bias as they did in Sugahara (2011). English listeners, on the other hand, showed no bias towards either direction. Although Seoul Korean listeners also showed an iambic bias, Japanese listeners' bias was significantly stronger than Korean listeners'. The results support the hypothesis that Japanese listeners' perception of English lexical stress is influenced by the unmarked antepenultimate accent pattern in their native language.

1. Introduction

Native English listeners are known to be biased towards word-initial stress when listening to English word stimuli (van Leyden & van Heuven, 1996; Cooper, Cutler & Wales, 2002), which has been considered to be due to the majority of the English vocabulary carrying strong word-initial syllables (Cutler & Carter, 1987; van Leyden & van Heuven, 1996; Cooper *et al*., 2002). In contrast, there is indication that Japanese listeners are biased towards non-initial stress when listening to English words (Sugahara, 2011). In Sugahara's (2011) perception study, I asked twelve native English listeners and thirty-eight Japanese listeners to judge the location of the syllable carrying primary stress in English disyllabic trochaic nouns (primary stress on the initial syllable and secondary stress on the final syllable) and disyllabic iambic verbs (secondary stress on the initial syllable and primary stress on the final syllable) whose segmental patterns were identical: *IMpact* vs. *imPACT*, *IMport* vs. *imPORT*, *REmake* vs. *reMAKE*, *REtake* vs. *reTAKE*, *TRANSplant* vs. *transPLANT*, *TRANSport* vs. *transPORT*. In Sugahara's (2011) experiment, those two-syllable words were produced by a female native speaker of American English and three types of stimuli were created out of them: the 'Natural' stimuli without any acoustic synthesis, the 'Flat' stimuli which were synthesized to have a flat F0 contour (all syllables were set at 175Hz), and the 'Declining' stimuli which were synthesized to have an F0 contour with 25Hz decline from the initial syllable (186Hz) to the second (161Hz). Both the native English and the Japanese listeners performed well when given the Natural stimuli, being able to distinguish the trochaic and the iambic forms correctly almost 90% of the time. In the Declining and the Flat stimuli, the trochaic and the iambic words were not disambiguated by pitch while other acoustic contrasts were kept intact. The F0 slope of the Declining contour, however, was so steep that the both groups of listeners gave initial stress responses about 70% to 80% of the time regardless of the original stress patterns of the stimuli. When given the Flat stimuli, however, the native English listeners preferred more initial stress than non-initial stress, i.e., 60 to 75% of their responses was of initial stress, regardless of the stress patterns of the original stimuli. In contrast, the Japanese listeners gave more non-initial stress responses, i.e., 70 to 77% of their responses was of final stress regardless of the stress patterns of the original stimuli. In summary, when F0 was not a reliable cue to hint the stress pattern of the English stimuli, the Japanese listeners showed a bias towards non-initial stress while the native English listeners showed a bias towards initial stress.

The question, then, is how to interpret the Japanese listeners' bias towards non-initial stress obtained in Sugahara (2011). One possible hypothesis is that they were influenced by the most unmarked accent pattern in Japanese loanwords, i.e., the antepenultimate accent. As will be discussed in Section 3.2, the two-syllable English nouns and verbs used in Sugahara's (2011) experiment both turn out to bear non-initial accent once they are converted into Japanese loanwords due to epenthetic vowel insertion after syllable-coda consonants and the application of the antepenultimate accent rule (see Section 3.2. for more details). It is possible to imagine that Japanese listeners hear an illusionary non-initial stress, being influenced by the antepenultimate accent rule in their L1, even when they are presented with L2 (English) stimuli. Under this hypothesis, it is expected that listeners whose L1 has neither lexical stress nor lexical accent will behave differently, in tasks similar to Sugahara's (2011), from both Japanese listeners who are biased towards non-initial stress and English listeners who are biased towards initial stress because there is nothing in their L1 to make those listeners biased towards either direction.

In order to test the hypothesis, a new auditory perception experiment was conducted in this study, in which not only English and Japanese listeners but also Seoul Korean (standard Korean) listeners participated. Seoul Korean listeners were included as participants this time because their L1 has no lexical stress/accent contrast. That is, the lexical prosody system of their L1 is expected to make them biased towards neither initial nor non-initial stress. If Seoul Korean listeners also exhibit a strong non-initial stress bias just as native Japanese listeners do when asked to locate English primary-stressed syllables in a perception experiment similar to that in Sugahara (2011), then there is no reason to conclude that native Japanese listeners' response pattern is influenced by the antepenultimate accent rule of loanwords in their L1.

The organization of this article is the following. In Section 2, I will give an overview of previous studies on the perception of L2 lexical prominence influenced by the lexical prosody systems of L1. Section 3 will introduce the lexical prosody systems of the three languages: English, Japanese and Seoul Korean. In the same section, some consideration to the hypothesis and its predictions will be given. Section 4 presents experimental procedures and results. In Section 5, discussion on the results and concluding remarks will be provided.

2. Previous studies on the perception of L2 lexical prominence influenced by L1 lexical prosody systems

Previous studies on the perception of L2 lexical prominence, i.e., lexical stress or accent, being influenced by L1 prosodic systems have been mainly focused on listeners' insensitivity or 'deafness' to suprasegmental information in L2 that does not contribute to lexical contrast in L1. One such example is Japanese listeners' perception of English lexical stress. Beckman (1986) has shown that Japanese listeners heavily rely on F0 and have little sensitivity to other cues such as duration and amplitude when discriminating disyllabic noun-verb pairs such as *DIgest* vs. *diGEST*, while native English listeners use multiple cues more or less evenly to discriminate them (also see Slujiter & van Heuven (1996) for native English listeners' usage of multiple acoustic cues to discriminate English stress patterns).[1] Japanese listeners' insensitivity to acoustic cues other than F0 is because their L1, i.e., Japanese, uses only F0 for accent realization, and their perceptional ability is fossilized to react only to the acoustic cues that are contrastive in their L1. (See Section 3 for more details about the nature of English stress and that of Japanese accent).

Another example is demonstrated by Duoux and his colleagues (Dupoux, Pallier, Sebastian & Mehler, 1997; Peperkamp & Dupoux, 2002, etc.). Dupoux *et al*. (1997) compared the perception of non-native stress by French listeners whose native language does not have contrastive stress with that by Spanish listeners whose native language has contrastive stress. In French, stress always falls on the final syllable of content words and there is no minimal pair of words that contrast in lexical stress locations. Spanish, on the other hand, allows minimal pairs that contrast in stress locations such as *tópo* ('mole') vs. *topó* ('met') though the most popular stress position in Spanish is the penultimate syllable. Dupoux *et al*. carried out ABX experiments with non-native nonsense words, and obtained a result that Spanish listeners outperformed French listeners when they were asked to distinguish the words only focusing on stress locations such as *bópelo* and *bopélo*. They also report that French listeners outperformed Spanish listeners when they were asked to ignore stress information to judge the identity of the non-native words. Although Cooper *et al*. do not reveal explicitly which acoustic cues were associated with the lexical stress in their nonsense word stimuli, it is likely that the stressed and the unstressed syllables were distinguished at least by F0. Assuming that this speculation is on the right track, their results show that Spanish listeners rely on F0 and possibly other acoustic cues when judging whether the words presented to them are identical or different even when those words are of non-native languages while French listeners tend to ignore those cues when doing the same task. Cooper *et al*. (1997) state that the suprasegmental cues associated with stress such as F0 are 'a non-detachable aspect of phonological information' (p.9) for Spanish listeners. Peperkamp and Dupoux (2002) use the term 'Stress Parameter' and claim that native listeners of languages with contrastive lexical stress like Spanish set their Stress Parameter early in the course of language acquisition such that stress-related suprasegmental information is encoded in their phonological representation. For French listeners, however, the Stress Parameter is set such that the information is not encoded in their phonological representation because it is not useful in their native language (Peperkamp & Dupoux, 2002). Peperkamp and Dupoux (2002) further report that Finnish and Hungarian listeners whose native languages lack lexical stress contrast behave in a similar way as French listeners.

I am, however, unaware of previous studies about the influence of lexical stress/accent distribution in L1 on the perception of lexical prominence in L2. It has been already shown that when native English listeners are presented with English stimuli with little or limited segmental information to discriminate different stress patterns, they rely on their stochastic knowledge of English lexical stress distribution to make a judgment on the stress pattern of the words they have heard (van Leyden & van Heuven, 1996; Cooper *et al*., 2002). Van Leyden and van Heuven carried out gating tasks, in which native English listeners were presented with not only English word stimuli with multiple syllables but also fragmental stimuli extracted from the original words, e.g., the stimuli with a word initial consonant only; those with a word initial syllable; etc. The original stimuli varied in terms of stress patterns: some had initial stress (Sw and Sww) while others had non-initial stress (wS, wSw, wwS). Listeners wrote down the words that they believed to be what they had heard each time they listened to each stimulus. Van Leyden and van Heuven report that the majority of error responses to the fragmental stimuli with a word-initial consonant and to those with a word-initial syllable were of initial stress answers regardless of the stress patterns of the original stimuli from which those fragments were extracted. In one of the experiments carried out by Cooper *et al*. (2002), native English listeners listened to the initial syllables of multi-syllable words forming minimal pairs in terms of stress locations, such as *MUsic* with initial tress and *muSEum* with non-initial stress. They were, then, asked to judge which member of the pair they had heard. Their correct response rate of initial-stress members was more than 70% while that of non-initial stress members were below 50%. That is, more than half of their responses to non-initial-stress stimuli were of initial stress. Van Leyden and van Heuven, and Cooper *et al*. both claim that those results indicate that native English listeners were influenced by the predominance of words with initial stress in their native vocabulary. It is not clear, however, whether listeners' knowledge of lexical prominence distribution in L1 even affects the perception of lexical prominence in L2, and the main goal of the current study is to clarify this point by comparing the perception of English lexical stress by Japanese listeners with that by native English listeners and Seoul Korean listeners in a forced identification experiment.

3. The lexical prosody systems of the three languages and predictions

An overview of the lexical prosody systems of the three languages, i.e., English, Japanese and Seoul Korean, is provided in Sections 3.1 to 3.3. At the same time, predictions derived from the hypothesis laid out in Section 1 are summarized in Section 3.4.

3.1. The phonetic and phonological characteristics of English lexical stress

English is known as a 'stress accent' language (McCawley, 1968; Beckman, 1986; among others), which is defined by both phonetic and distributional terms. Phonetically speaking, stress accent languages are those whose lexical prominence is associated with multiple acoustic parameters, for example vowel quality, pitch (fundamental frequency: F0), intensity, duration, and spectral balance (Beckman, 1986; Beckman & Edwards, 1994; Sluijter & van Heuven, 1996; Okobi, 2006; Plag, Kunter & Schramm, 2012; among others). Distributionally speaking, Hyman (2006) proposes that stress accent languages be those that satisfy the following two criteria: 'obligatoriness', i.e., every lexical word has at least one primary-stressed syllable, and 'culminativity', i.e., every lexical word has at most one primary-stressed syllable. English satisfies those two criteria. English is also known as a 'free stress system' language in which the location of lexical stress varies from word to word. Nonetheless, there is a strong tendency that the language prefers word-initial primary stress. In the following part of this section, the phonetic nature and the distributional characteristics of English stress are introduced.

3.1.1. The phonetic nature of English lexical stress

As already mentioned above, every lexical word in English obligatorily has one and only one primary-stressed syllable. Furthermore, they may optionally have one or more than one secondary-stressed syllable. Regardless of the difference between the two stress levels, English stressed syllables always have a full vowel (Selkirk, 1980; Beckman & Edwards, 1994; among others). Unstressed syllables, on the other hand, have a reduced vowel, e.g., schwa [ə], in principle.[2] Therefore, the presence or absence of stress results in a vowel quality difference as seen in the comparison of *Jaˈpan* [ʤə.ˈpʰæn] vs. *ˌJapaˈnese* [ˌʤæ.pə.ˈniz], where the unstressed reduced vowel [ə] and the full vowel [æ] with primary stress in the former contrast with the full vowel [æ] with secondary stress and the unstressed reduced vowel [ə] in the latter respectively.

Although both primary-stressed and secondary-stressed syllables carry full vowels, they are different in terms of (a) co-occurability with a ‘nuclear pitch accent’, which ultimately results in differences in F0 and overall intensity, and (b) duration and spectral balance. Let us consider the first point: it is only primary-stressed syllables that may co-occur with a nuclear accent (Beckman & Edwards, 1994; Shattuck-Hufnagel, 1995). The term ‘pitch accent’ was originally proposed by Bolinger (1958), which is defined as a pitch contour or a tonal melody that co-occurs with a prominent syllable and often involves an F0 peak or trough (Ladd, 2008). A nuclear pitch accent, then, is the rightmost accent in an intonational phrase, which is the most prominent accent of all in the same phrase (Pierrehumnert, 1980; Beckman, 1986).[3] The conditions that determine which lexical word in a sentence should bear a nuclear pitch accent are relatively complicated as multiple factors such as semantic, pragmatic, syntactic and prosodic factors come into play (Gussenhoven, 1984; Selkirk, 1984, 1995; Terken & Hirschberg, 1994; among others). What is for sure is that a lexical word produced in isolation obligatorily has a nuclear pitch accent on its primary-stressed syllable (Gussenhoven & Bruce, 1999; Gussenhoven, 2004; among others). It is also true that a lexical word which is 'narrowly focused' also bears a nuclear accent. In contrast, words interpreted as already given in a discourse that follow a narrowly focused word are usually deaccented. A word is narrowly focused if it is the only focused word within a given linguistic domain (Ladd, 1980, 2008). For example, consider the sequence of a question-answer pair, Q: *Who visited us?* A: *Annabel visited us*. Here, the proposition [someone visited us] has been already presupposed by conversation participants, and the function of the answer sentence is to identify who the person that visited is, i.e., *Annabel*. In this case, *Annabel* is narrowly focused, and the word bears a nuclear pitch accent followed by words that are interpreted as already given, i.e., *visited us*, which are all accentless. I will call those given items that come after a narrowly focused word 'post-focus given words' henceforth. Within the word ˈ*Anna*ˌ*bel*, the initial syllable with primary stress ˈ*An* bears a nuclear pitch accent, and the final syllable with secondary stress ˌ*bel* is accentless. When a broader domain such as the entire sentence or phrase is focused, the last lexical word of the domain generally bears a nuclear pitch accent (Ladd, 1980; Halliday, 1967).[4] This is the case of what Ladd (1980, 2008) calls 'broad focus', or 'normal stress' as has been traditionally called. For example, when being asked *What happened?* and answering *I just saw ˈAnnaˌbel*, the entire answer sentence forms a focus domain. In this case, it is the final lexical word *ˈAnnaˌbel* that bears a pitch accent, and here, too, the accent goes to the primary stress of the word.

Although pitch accents in English may take various tonal shapes, the most frequently observed ones are those which involve a high accent (H*) tone (Dainora, 2006).[5] According to Dainora, pitch accents that involve an H* accent tone account for 95% of the whole occurrences of pitch accents in English. In other words, English primary-stressed syllables with a pitch accent are realized with high F0 peaks most of the time. Now consider the verb-noun pair of *transplant* /træns.plænt/, one of the disyllabic word pairs that were used in Sugahara's (2011) experiment. The word *transplant* consists of two stressed syllables: both the initial and the final syllable contain a full vowel /æ/. When it is produced as a verb, the initial syllable bears secondary stress and the final syllable primary stress. When the verb is produced in a context in which it bears a nuclear pitch accent, the final primary-stressed syllable /ˈplænt/ coincides with a nuclear pitch accent, while the initial secondary stress syllable /træns/ could be pitch-accentless and realized with lower pitch. This is shown in the pitch track of Figure 1, which was obtained from female speech used in Sugahara (2012).

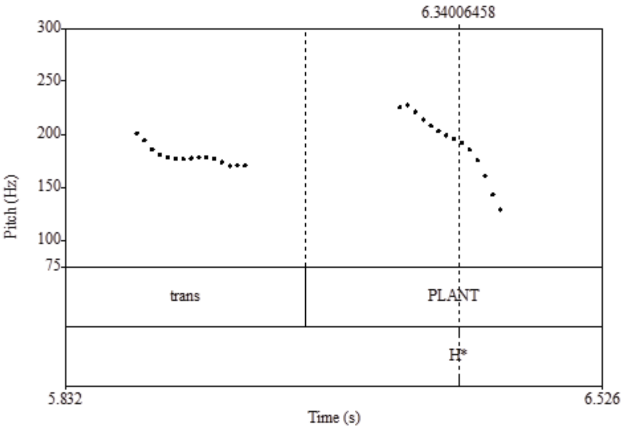

Figure 1. The pitch track of *transplant* (verb) produced with a nuclear pitch accent. The final syllable with primary stress is capitalized.

The noun counterpart with initial primary stress and final secondary stress /ˈtræns.ˌplænt/, on the other hand, shows the opposite pattern: the nuclear pitch accent appears at the initial syllable as shown in Figure 2.

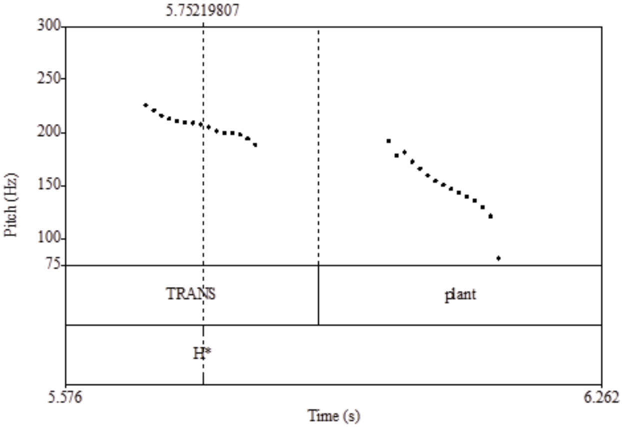

Figure 2. The pitch track of *transplant* (noun) produced with a nuclear pitch accent. The initial syllable with primary stress is capitalized.

When words are made into post-focus given words, not only secondary-stressed syllables but also primary-stressed syllables lack a pitch accent. As a result, in the post-focus part of an utterance, both the verb form and the noun form of *transplant* are realized with a flat pitch contour without conspicuous pitch rise anywhere. This is shown in Figures 3 and 4.

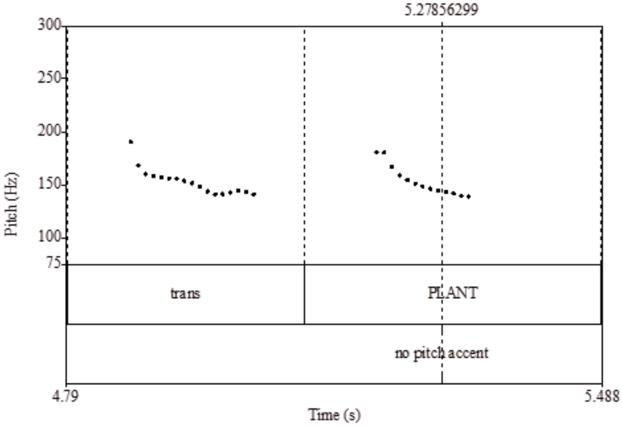

Figure 3. The pitch track of *transplant* (verb) produced in a post-focus (accentless) position. The final syllable with primary stress is capitalized.

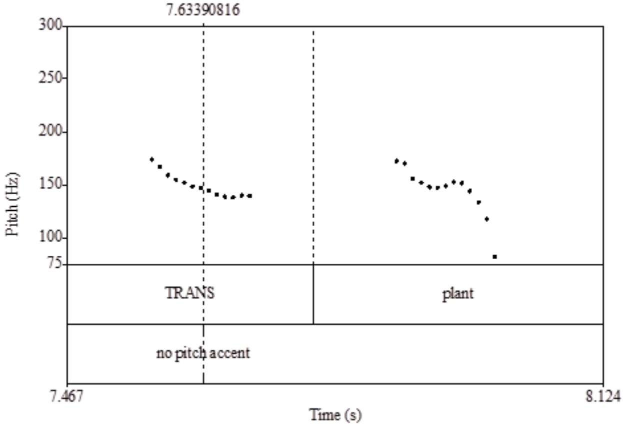

Figure 4. The pitch track of *transplant* (noun) produced in a post-focus (accentless) position. The initial syllable with primary stress is capitalized.[6]

In summary, one of the important differences between primary and secondary stress in English is that primary-stressed syllables may co-occur with a nuclear pitch accent while secondary-stressed syllables may not. Pitch accents in English involve an H* accent tone most of the time, and as a result primary-stressed syllables with a pitch accent are usually higher in F0 than secondary-stressed syllables. This point has been already reported by many researchers such as Beckman (1986), Fujisaki, Hirose & Sugito (1986), Sluijter & van Heuven (1996), Plag *et al*. (2011), Sugahara (2012), among others. Another acoustic parameter closely correlated with F0 is overall intensity. Overall intensity gets greater as F0 gets higher. It is because in order to be heard as having the same loudness, high-pitched sounds should have greater intensity than low-pitched tones. Therefore, the pitch-accented syllables with primary stress usually have greater intensity than the syllables with secondary stress (Beckman, 1986, Sluijter & van Heuven, 1996; Plag *et al*., 2011; among others).

In addition to F0 and overall intensity, duration and spectral balance, i.e., the difference between the amplitude of a higher frequency band and that of a lower frequency band, are different between primary-stressed syllables and secondary-stressed syllables. Sluijter & van Heuven (1996), de Jong (2004), Okobi (2006) and Sugahara (2012) report that primary-stressed syllables with a nuclear pitch accent are longer than secondary-stressed syllables. According to Sluijter, Shattuck-Hufnagel, Stevens & van Heuven (1995) and Okobi (2006), pitch-accented primary-stressed vowels have greater amplitude in a higher frequency band than secondary-stressed syllables, which means that the difference between the amplitude of a lower frequency band and that of a higher frequency band is smaller in pitch-accented primary-stressed vowels than in secondary-stressed syllables.

That is, the spectral energy distribution is more balanced in pitch-accented primary-stressed vowels than in secondary-stressed vowels (Sluijter et al. 1995 and Okobi 2006). It has been further reported that primary-stressed and secondary-stressed vowels are differentiated in terms of duration and spectral balance even in accentless environments such as in the post-focus part of an utterance (for duration, see Sluijter & Heuven, 1996; de Jong, 2004; Okobi, 2006; Sugahara, 2012; for spectral balance, see Sluijter *et al.*, 1995; Okobi, 2006; Plag *et al*. 2011).

3.1.2. The distributional characteristics of English stress

English is a free stress-system language in which stress locations vary from word to word, which sometimes results in semantic contrast between two lexical items of the same grammatical category. For example, in spite of the fact that *differ* and *defer* are both verbs and have the same segmental content [dɪfɝ] in many English dialects, the former has primary stress on the initial syllable and the latter on the final syllable. Nonetheless, there is a strong tendency that disyllabic and polysyllabic words, especially nouns, prefer penultimate or antepenultimate primary stress (Carr, 2012). As a result, bisyllablic and trisyllabic words tend to have word-initial primary stress. Needless to say, all monosyllabic lexical words carry word-initial stress. Cutler and Carter (1987) looked into the MRC Psycholinguistic Database (Coltheart, 1981), which contains 33,313 phonetically transcribed words, and counted the number of lexical words that begin with strong syllables. According to them, 33,060 words among those are lexical words, and 62% of the lexical words have word-initial primary stress (see Cutler & Carter's Tables III). They further analyzed the frequency count of words that appeared in a spontaneous speech corpus (*London-Lund Corpus of*

English Conversation by Svartvik & Quirk, 1980), and showed that lexical tokens with word-initial primary stress (67,424 tokens) accounted for 88% of the whole lexical tokens (76,963 tokens) in the corpus (see Cutler & Carter's Table V). Given this word-initial primary stress predominance, it is no wonder that English native speakers were biased towards word-initial primary stress in not only van Leyden and van Heuven (1996) and Cooper *et al.* (2002) but also Sugahara (2011).

3.2. The Japanese lexical accent system

Unlike English, a stress accent language, Japanese is often referred to as a 'pitch accent language' (McCawley, 1968) or 'non-stress accent language' (Beckman, 1986). Non-stress pitch accent languages are those which utilize only pitch to realize lexical prominence contra stress accent languages that use other acoustic parameters in addition to pitch for lexical prominence (Beckman, 1986). In most of the Japanese dialects including the major ones such as Tokyo and Kyoto-Osaka Japanese, a pitch accent is always realized as an abrupt pitch fall from an accented mora to the following mora and is phonologically represented as a bitonal accent tone H*+L.

Both Tokyo and Kyoto-Osaka Japanese do not satisfy the 'obligatoriness' constraint: they allow both accented and unaccented words, and the presence or absence of accent contributes to meaning differences, e.g., *ha˥shi* 'chopsticks' vs. *hashi* 'an edge' in Tokyo Japanese; *ha˥shi* 'a bridge' vs. *hashi* 'an edge' in Kyoto-Osaka Japanese.[7] According to Kubozono's (2006) investigation on three-mora nouns in Tokyo Japanese, only about 30% of native words are accented and the rest are unaccented. When it comes to Sino-Japanese words and loanwords in Tokyo Japanese, the percentage of accented words amounts to 43% and 93% respectively (Kubozono, 2006).

Kyoto-Osaka Japanese also shows a similar tendency.[8] I looked into *Osaka/ Tokyo akusento onsei ziten* (*OTAOZ*) (Pronunciation dictionary of Osaka/ Tokyo accent) by Sugito (1995) to see the distribution of accented and unaccented nouns belonging to each of the three lexical groups in Osaka Japanese. The dictionary contains the accent patterns produced by six native speakers of Osaka Japanese (three elderly people born in between 1916 and 1932, and three younger people born in the early 1960's). I made a count of accented nouns in each lexical group in the dictionary.[9] When doing so, I only extracted nouns whose accent patterns were agreed on by all the six speakers. According to my count, 15% of native Japanese words, 55% of Sino-Japanese words and 98% of loanwords are accented in Osaka Japanese. Those figures are summarized in Table 1. What is clear from the table is that, in both Tokyo and (Kyoto-)Osaka Japanese, three-mora loanwords are almost homogeneously accented, which means that when Japanese native speakers introduce a new word from English into the Japanese lexicon, an accent is almost always assigned to it.

Table 1. The percentage of accented nouns

Word type	Tokyo Japanese (from Kubozono, 2006)	Osaka Japanese
Native	29%	15% (158/1,030)
Sino Japanese	49%	55% (1,799/3,280)
Loanwords	93%	98% (680/694)

Now consider the distribution of accents in accented words only. Accent locations vary from word to word, which contribute to lexical contrast, e.g., *ha˥shi* 'chopsticks' vs. *hashi˥* 'a bridge' in Tokyo Japanese. Furthermore, any syllable in a word can potentially be a possible site of accent assignment, e.g., *ka˥.ra.su* 'crow', *ko.ko˥.ro* 'mind', *a.si.ta˥* 'tomorrow' in Tokyo Japanese. Having said that, not all possible accent locations are equally popular. Among three-mora accented nouns in Tokyo Japanese, those which carry an accent on the antepenultimate mora are the commonest in every vocabulary stratum (Kubozono, 2006) as shown in Table 2. (Kyoto-)Osaka Japanese shows similar distribution. As I looked into *OTAOZ* by Sugito (1995), the percentage of three-mora Osaka Japanese nouns in each lexical group with the antepenultimate accent is 54% in the native vocabulary, 97% in Sino Japanese, and 97% in loanwords (see Table 2).[10] For example, English words with penultimate stress such as *casíno* and *pajáma* come to have an accent on the antepenultimate mora (*ka˥.zi.no* and *pa˥.ja.ma*) once they are introduced into loanwords not only in Tokyo Japanese but also in (Kyoto-) Osaka Japanese.[11]

Table 2. The percentage of nouns with antepenultimate accent among all accented nouns

Word type	Tokyo Japanese (from Kubozono, 2006)	Osaka Japanese
Native	59%	54% (85/158)
Sino Japanese	95%	97% (1,746/1,799)
Loanwords	96%	97% (660/680)

The antepenultimate accent is favored in longer words, too. According to Tanaka (2009), 40 to 50% of accented loanwords with five morae have an accent on the antepenultimate position in both Tokyo Japanese and Osaka Japanese, and the percentage is greater than that of any other accent locations: penultimate accent is almost none, pre-antepenultimate accent is about 25 to 27%, and initial accent is 18 to 24%. The dominance of the antepenultimate accent has led researchers to propose the rule known as the ‘antepenultimate rule’ which states that the default accent location is the syllable that contains the antepenultimate mora (Akinaga, 1981; Shibata, 1994; Kubozono, 2006; among others).

One thing to notice here is that the Japanese antepenultimate rule could be the source of the Japanese listeners’ strong preference for final stress when asked to detect the location of English stress in Sugahara’s (2011) perception experiment in which pitch information was not a reliable cue for the location of stress. As already mentioned in Section 1, the English stimuli used in the experiment were of two-syllable nouns and verbs that shared exactly the same segmental content but differed in the location of stress: nouns with primary stress on their initial syllables and verbs with primary stress on their final syllables (*IMpact*(n) vs. *imPACT*(v), *IMport*(n) vs. *imPORT*(v), *REmake*(n) vs. *reMAKE*(v), *REtake*(n) vs. *reTAKE*(v), *TRANSplant*(n) vs. *transPLANT*(v), *TRANSport*(n) vs. *transPORT*(v)). The key point here is that their final syllables are all super-heavy consisting of three morae, and the antepenultimate morae in those words correspond to the nucleus vowels of the final syllables. A possible scenario is that the Japanese participants in Sugahara’s (2011) experiment were strongly affected by the antepenultimate rule in Japanese and had an assumption that the antepenultimate morae, i.e., the final syllable nuclei, were the location of stress even in English.

Or it is even possible to imagine a situation that the native Japanese listeners converted those English stimuli into the pronunciation of corresponding Japanese loanwords in their heads, and as a result they heard an illusionary accent on the antepenultimate mora. The final syllable nucleus vowels of the original English words correspond to the antepenultimate mora in the Japanese loanword counterparts in spite of the fact that the number of syllables in those two forms are different. When English words with consonant clusters and coda consonants that are phonotactically illegal in Japanese are adopted into Japanese as loanwords, there is an increase in the number of syllables due to vowel epenthesis to fix the shape of the illegal segmental sequences. It is true for the English disyllabic words used in Sugahara (2011), too. Although it increases the number of syllables, vowel epenthesis in coda-consonant positions does not change the mora count of the original English forms. Therefore, the number of the morae after the final-syllable nucleus vowel in the English disyllabic words used in Sugahara (2011) is unchanged even after they are adopted into Japanese with vowel epenthesis as shown in (1), and the antepenultimate morae of the Japanese loanwords correspond to the nucleus vowels of the final syllables in the original English words. (The periods in the English examples represent syllable boundaries while those in the Japanese examples show mora boundaries. The number above each mora is the count from the final mora.)

(1)

a. English stimuli b. Corresponding Japanese loanwords

3 2 1 3 2 1 → 3 2 1 3 2 1
_IM.**pa**ct, im.**PA**CT_ → _i⌝.m.**pa**.k<u>.t<o> or i.m.**pa**⌝.k<u>.t<o>_

3 2 1 3 2 1 → 3 2 1
_IM.**po**rt, im.**PO**RT_ → _i.m.**po**⌝.o.t<o>_

32 1 32 1 → 3 2 1
_RE.**ma**ke, re.**MA**KE,_ → _ri.**me**⌝.i.k<u>_

32 1 32 1 → 3 2 1
_RE.**ta**ke, re.**TA**KE_ → _ri.**te**⌝.i.k<u>_

3 2 1 3 2 1 → 3 2 1
TRANS.p**la**nt, trans.P**LA**NT → _t<o>.ra.n.s<u>.p<u>.**ra**⌝.n.t<o>_

3 2 1 3 2 1 → 3 2 1
TRANS.**po**rt, trans.**PO**RT → _t<o>.ra.n.s<u>.**po**⌝.o.t<o>_

Let us consider the above-mentioned more concretely. All of the original English words in (1a) end with a coda consonant or consonants. Since Japanese allows only CV syllables in principle with the exceptions of the nasal coda and the first part of geminates, vowel epenthesis takes place after the illegal coda consonants to fix the syllable structure when those words in (1a) are adopted into Japanese loanwords as in (1b). In (1b), epenthetic vowels are bracketed by ‘< >’. Another thing is that the rhotic consonant /r/ in the coda position of the original English word such as *import* is converted into the latter half of a long vowel in the Japanese loanword counterpart as in *impoot*<o>. The antepenultimate mora of the Japanese loanwords and the corresponding nucleus vowel of the final syllables in the original English words are displayed with bold-faced letters in (1). The bold-faced moras in the Japanese loanwords are also where an accent falls. The only exception is *i.m.pa.k<u>.t<o>* ‘impact’: in Tokyo Japanese, the initial mora but not the bold-faced antepenultimate mora is accented. In Osaka Japanese, however, both initial and antepenultimate accent are possible for the same word according to *OTAOZ* (Sugito, 1995).

I have laid out in this section that the antepenultimate position is the default accent location in Japanese loanwords, which corresponds to the final syllable nuclei of the source English words with super heavy final syllables. This is exactly the case of the disyllabic words used in Sugahara (2011), and the speculation that the Japanese listeners’ perception of English lexical stress in those words might be influenced by the Japanese antepenultimate accent is not far-fetched.

3.3. The Korean lexical prosody systems

Most of the varieties of Korean, unlike Tokyo and Kyoto-Osaka Japanese, lack a lexically specified accent, and Seoul Korean, i.e., the standard dialect, is not an exception (Kim, 2013).[12] It does not mean that F0 ups and downs are absent in the language because tones appear to mark phrase edges and word-initial segmental contrast (Jun, 1998, 2005, 2006). According to Jun, when a minor phonological phrase consists of more than three syllables[13], the initial two syllables and the final two syllables are associated with edge tones: the latter are always associated with an LH (rising) contour unless they are at the final position of an intonation phrase.[14] The first two syllables of a minor phonological phrase are associated with an HH contour if the onset consonant of the initial syllable is 'aspirated' or 'fortis (tense)', and they are associated with an LH rising contour otherwise (Jun, 1998)[15]. That is, when a minor phonological phrase is four-syllable long or longer, the phrase is associated with the tonal sequence of either LH...LH or HH...LH. When it is less than four-syllable long, however, not all of the edge tones are phonetically realized due to 'undershoot', i.e., the length of the minor phonological phrase is too short to phonetically realize all of the four tones. It is always the middle tones of the tonal sequence that undergo undershoot, i.e., L(H L)H, L(H)LH, LH(L)H, H(H L)H, H(H)LH, and HH(L)H, where the tones in the parentheses are the ones that undergo undershoot (Jun, 2005, 2006). In this way, there are tones in Seoul Korean speech, too, which are regulated by linguistic factors such as minor phonological phrase formation and segmental contrasts. The tones, however, have nothing to do with lexical prominence such as accent or stress.

3.4. Predictions

I gave an overview of the lexical prosody systems of the three languages: English, Japanese, and Seoul Korean. Since the majority of English syllables with primary stress occupy the initial position of lexical words, it is no wonder that native English listeners exhibited strong preference for initial stress not only in van Leyden and van Heuven (1996), and Cooper *et al.*

(2002) but also in Sugahara (2011). That native listeners are influenced by their knowledge of lexical stress distribution in their L1 in perception tasks with L1 stimuli is not surprising at all.

The question, then, is whether or not Japanese listeners' strong preference for final stress when hearing English disyllabic words is due to the antepenultimate rule in their L1. The most unmarked position of Japanese lexical accent, i.e., the antepenultimate mora, corresponded to the final syllable nucleus vowel of the English stimuli in Sugahara's (2011) experiment. Given this, a possible hypothesis is that the antepenultimate rule in their L1 is the source of the strong bias towards non-initial (final) stress exhibited by the Japanese listeners in Sugahara's (2011) study. Another possibility is that the Japanese listeners' strong preference for final stress is just a default or universal reaction by non-native listeners who have not fully acquired the statistic knowledge that initial stress is the commonest in the English vocabulary.

To prove the hypothesis that the antepenultimate rule in Japanese is the source of Japanese listener's strong preference for non-initial stress, Japanese listeners were compared with native English listeners and Seoul Korean listeners in the current study. If the hypothesis is on the right track, Seoul Korean listeners are predicted to behave differently from both native English listeners and Japanese listeners because there is neither lexical stress nor lexical accent in their L1 that affects their perception of English stress. That is, they are predicted to exhibit no bias or a less strong bias towards either initial or final stress in an experiment similar to Sugahara's (2011), and their responses should be somewhere in between those by English listeners and those by Japanese listeners as summarized in Table 3.

Table 3. The summary of predictions

	English listeners (native listeners)	Seoul Korean listeners (non-native listeners)	Japanese listeners (non-native listeners)
Predicted results	strongly biased towards initial stress	not or less biased towards either direction	strongly biased towards final stress
Source	influenced by the predominance of initial stress in the English vocabulary.	no influence from L1 lexical stress/accent because there is no stress/accent contrast in L1.	influenced by the antepenultimate accent rule (or the predominance of antepenultimate accent) in the L1 vocabulary.

4. Experiment

A forced choice identification task was carried out to test the hypothesis mentioned above, of which details and results are shown below.

4.1. Materials

4.1.1. Words used as stimuli

The stimuli used in this experiment included not only two-syllable words (e.g., *TRANSplant* and *transPLANT*) but also three-syllable ones (e.g., *INtercept* and i*nterCEPT)*. Those two-syllable and three-syllable words are shown in (2) and (3) respectively.

(2) Two-syllable words

	Nouns (Initial Primary Stress)	Verbs (Final Primary Stress)
a.	*IMport*	*imPORT*
b.	*INsult*	*inSULT*
c.	*MISprint*	*misPRINT*
d.	*REtake*	*reTAKE*
e.	*TRANSplant*	*transPLANT*

(3) Three-syllable words

	Nouns (Initial Primary Stress)	Verbs (Final Primary Stress)
a.	*INtercept*	*interCEPT*
b.	*OVerthrow*	*overTHROW*
c.	*UNdershoot*	*underSHOOT*

The pair of *overthrow* does not have an antepenultimate accent when its noun and verb are adopted into Japanese as a loanword: both are pronounced as *o.o.va.a.s<u>.ro⌝.u*, with an accent on the penultimate mora. It is not problematic to the hypothesis suggested in this study because the antepenultimate accent rule of Japanese loanwords allows exceptions: when the antepenultimate mora contains an epenthetic vowel and the final syllable is heavy, the penultimate mora is accented (Kubozono, 2001; Kawahara, 2015), as in *t<u>.ri⌝.i* 'tree', *s<u>.ri⌝.i* 'three', *s<u>.ro⌝.o*. That is, the exceptional penultimate accent here is rule-governed and is also part of the 'larger' loanword accent rule in Japanese. It should be further noted that the penultimate mora with the exceptional accent also corresponds to the nucleus vowel of the final syllable in the original English word.

4.1.2. Recording procedure

The words in (2) and (3) were produced by a male American English speaker in his mid-twenties, and recorded onto Marantz Solid State Recorder PMD671 (44.1 KHz, 16 bits), using Countryman ISOMAX Headset Microphone, in a sound-attenuated room. Each word was embedded in a sentence "*I wanted to say* ___" when words started with a consonant and in a sentence "*I said* ___" when words started with a vowel or a liquid 'r'.

Those sentences were presented on a computer monitor, which the speaker read aloud. The location of primary stress was indicated by an accent mark ‘′’ placed on the vowel with primary stress so that the speaker could tell whether the word is trochaic or iambic. The speaker always placed an H* nuclear pitch accent on the primary-stressed syllables of those target words.

4.1.3. The acoustic nature of the words produced by the speaker

The wave forms of those target words were read into Praat and segmented from the carrier sentences. The vowel intervals and the foot intervals of those words were further segmented from the target words. Sonorant codas such as /r/ and /n/ in the final syllables of *import*, *insult*, *misprint*, *transplant* were included in the vowel intervals because boundaries between the vowels and the following sonorant codas were unclear. However, the vowel intervals of the initial syllables as in *import*, *insult*, *transplant*, *intercept* were separated from the following sonorant codas because their boundaries were clear. Another thing is that the onset consonant /r/ and the vowel in the initial syllable of *retake* was inseparable, and both the preceding sonorant and the following vowel were segmented together not only as a foot interval but also as a single vowel interval. For the foot intervals, their durations were measured. For the vowel intervals, not only durations but also peak F0 (pitch) values, overall intensities and spectral balance were obtained. In addition to those parameters that are said to disambiguate different stress patterns, formant values (F1 and F2) were also extracted for clarification purposes.

Peak F0 values, durations and overall intensities of the vowel intervals are shown in Table 4. Overall intensity values are those averaged over each of the vowel intervals. The durations of foot intervals are shown in Table

5. Spectral balance was obtained following Plag *et al*'s (2011) procedure[16]: a long-term average spectrum of each vowel interval was divided into two frequency bands (the lower band 0 to 1,000Hz and the higher band 1,000 to 4,000Hz), and intensities averaged over the lower frequency band was subtracted from those averaged over the higher frequency band ($B = I_{high}-I_{low}$). Since the higher frequencies are generally associated with lower amplitude than the lower frequencies, the spectral balance value *B* is expected to be negative. The *B* values of vowels with primary stress are shown in Table 6. F1 and F2 values were extracted from the mid-point of each vowel interval, and are provided in Table 7.

Table 4. The F0 peak, the duration and the overall intensity of each of the vowel intervals. The initial syllable vowels and the final syllable vowels are denoted as 'v_i' and 'v_f' respectively. The index 'v_i- v_f' means the difference between the relevant acoustic value of v_i and that of v_f.

Word	Category	Stress Pattern	F0 Peak (Hz)			Duration (sec)			overall Intensity (dB)		
			v_i	v_f	v_i- v_f	v_i	v_f	v_i- v_f	v_i	v_f	v_i- v_f
import	N	1 2	125.4	93.9	31.5	0.067	0.168	-0.101	72	59.1	12.9
	V	2 1	88.2	117.9	-29.7	0.052	0.184	-0.132	63	60.2	2.8
insult	N	1 2	107.5	74	33.5	0.087	0.173	-0.086	67	53.6	13.4
	V	2 1	92.7	105.2	-12.5	0.059	0.217	-0.158	64	56.7	7.3
misprint	N	1 2	100.7	96.3	4.4	0.069	0.18	-0.111	68.8	58.3	10.5
	V	2 1	85.6	99.7	-14.1	0.062	0.2	-0.138	61.9	60.2	1.7
retake	N	1 2	114.1	109	5.1	0.151	0.172	-0.021	63	57.4	5.6
	V	2 1	92.5	114.8	-22.3	0.13	0.169	-0.039	59	60.5	-1.5
transplant	N	1 2	110.5	89.7	20.8	0.126	0.196	-0.07	67	57.1	9.9
	V	2 1	92.2	108	-15.8	0.082	0.208	-0.126	63.4	62	1.4
intercept	N	1 2	111.6	83.5	28.1	0.054	0.145	-0.091	70.5	55.7	14.8
	V	2 1	94.1	95.1	-1	0.038	0.141	-0.103	61.8	61.3	0.5
overthrow	N	1 2	111.4	83.8	27.6	0.112	0.314	-0.202	63.9	52.1	11.8
	V	2 1	95.4	101.4	-6	0.096	0.357	-0.261	61	58.5	2.5
undershoot	N	1 2	109.9	84.4	25.5	0.091	0.165	-0.074	66.6	53.7	12.9
	V	2 1	93.1	105.5	-12.4	0.08	0.164	-0.084	65.8	55.8	10

Table 5. The foot interval durations

Word	Category	Stress Pattern	Duration (sec)		
			foot$_i$	foot$_f$	f_i-f_f
import	N	1 2	0.145	0.508	-0.363
	V	2 1	0.131	0.575	-0.444
insult	N	1 2	0.187	0.501	-0.314
	V	2 1	0.156	0.585	-0.429
misprint	N	1 2	0.22	0.454	-0.234
	V	2 1	0.182	0.524	-0.342
retake	N	1 2	0.151	0.431	-0.279
	V	2 1	0.13	0.456	-0.326
transplant	N	1 2	0.329	0.481	-0.151
	V	2 1	0.264	0.506	-0.242
intercept	N	1 2	0.246	0.487	-0.241
	V	2 1	0.202	0.553	-0.351
overthrow	N	1 2	0.252	0.446	-0.194
	V	2 1	0.236	0.506	-0.271
undershoot	N	1 2	0.263	0.554	-0.292
	V	2 1	0.266	0.556	-0.29

Table 6. Spectral balance

Word	Category	Stress Pattern	Spectral Balance (dB) ($B = I_{high}$-I_{low})		
			v_i	v_f	v_i- v_f
import	N	1 2	-22.6	-14.2	-8.4
	V	2 1	-22.1	-15.4	-6.7
insult	N	1 2	-24.2	-24.8	0.6
	V	2 1	-26	-19.5	-6.5
misprint	N	1 2	-22.9	-19.9	-3
	V	2 1	-25.5	-20.1	-5.4
retake	N	1 2	-10.9	-11.7	0.8
	V	2 1	-9.5	-11.8	2.3
transplant	N	1 2	-16.4	-20.2	3.8
	V	2 1	-17.9	-19.2	1.3
intercept	N	1 2	-24.5	-16.4	-8.1
	V	2 1	-23.6	-12.6	-11
overthrow	N	1 2	-17.4	-17.3	-0.1
	V	2 1	-22.8	-10.9	-11.9
undershoot	N	1 2	-10	-19.1	9.1
	V	2 1	-12.4	-15.7	3.3

Table 7. Formant values at the vowel midpoint

Word	Cateory	Stress Pattern	v_i		v_f	
			F1(Hz)	F2(Hz)	F1(Hz)	F2(Hz)
import	N	1 2	500	2517	583	963
	V	2 1	597	2071	640	903
insult	N	1 2	457	2439	676	737
	V	2 1	543	2209	604	886
misprint	N	1 2	470	1843	575	1813
	V	2 1	500	1036	610	1876
retake	N	1 2	354	1702	502	1978
	V	2 1	369	1629	468	2068
transplant	N	1 2	659	1464	631	1750
	V	2 1	644	1644	641	1771
intercept	N	1 2	519	2356	616	1557
	V	2 1	447	1652	668	1547
overthrow	N	1 2	435	954	581	1373
	V	2 1	524	852	643	1350
undershoot	N	1 2	666	1171	407	1944
	V	2 1	657	1178	396	1956

From Table 4, it is clear that all noun-verb pairs were distinguished by the F0 peak values: v_i had a higher F0 peak value than v_f in the trochaic noun forms while the relationship was the other way round in the iambic verb forms. As for the overall intensity, the difference between v_i and v_f was greater in the trochaic nouns than in the iambic verbs. As for the vowel and the foot durations, too, the relationship of the initial and the final positions was different between the trochaic nouns and the iambic verbs: the iambic verbs had greater durational difference between the two positions than the trochaic nouns in almost all the pairs. The only exception was the pair of *undershoot*. As for the spectral balance in Table 6, there seems to be no consistent difference between the trochaic nouns and the iambic verbs. In summary, the trochaic noun stimuli and the iambic verb stimuli had different relationship between the initial and the final vowels (or feet) in terms of durations and overall intensities in addition to F0 peak values.

4.1.4. Synthesis

Three F0 contours were prepared for each word: 'Natural', 'Flat' and 'Slightly Declining'. The Natural contour is of the original sounds produced by the speaker without any pitch manipulation, in which F0 cues to distinguish the two stress patterns were kept intact. Stimuli with the Flat contour and those with the Slightly Declining contour were created by modifying the F0 of the original sounds via the pitch manipulation function of Praat. In those synthesized stimuli, the left edge and the right edge of the initial and the final vowels were connected by a straight F0 line. That is, both the Flat and the Slightly Declining stimuli had a 'straight' pitch contour. In that sense, those two types of stimuli were very similar. In English natural speech, more or less flat F0 contours appear in the post-focus part of an utterance, where words are interpreted as given and pitch accents are absent (see Section 3.1), and the F0 contours of the Flat and the Slightly Declining contours in this experiment were created after the F0 pattern of post-focus given words. The assignment of F0 values to the Flat and the Slightly Declining contours followed the following process. First, the F0 ratio of the vowels in pitch-accented new words and the vowels in post-focus given words were estimated based on the male speech data provided in Sugahara (2012, Tables 7.1 and 7.2, p.215). The mean F0 values of the accentless post-focus primary-stressed vowels was about 83 to 85% of that of the pitch-accented primary-stressed vowels in the same word position, and the F0 of the post-focus secondary-stressed vowels was about 91 to 100% of that of the non-post-focus vowels in the same word position. Secondly, the F0 values of the primary-stressed vowels and those of the secondary-stressed vowels in the Natural stimuli of the current study were each averaged across the two positions (initial and final) and across all the eight words. Then, the F0 ratio of the post-focus and the pitch-accented primary-stressed vowels was applied to the averaged F0 value of the primary-stressed vowels in the Natural stimuli. In the same way, the F0 ratio of the post-focus and the non-post-focus secondary-stressed vowels was applied to the averaged F0 value of the secondary-stressed vowels in the current Natural stimuli. The outcome was about 90Hz for the primary-stressed vowels and about 86Hz for the secondary-stressed vowels. That is, those are the F0 values of the primary- and the secondary-stressed vowels estimated for the current stimuli if they are to be produced in the post-focus part of an utterance. Finally, the F0 contours of the synthesized stimuli, i.e., both the Flat and the Slightly Declining stimuli, were made to fall in the pitch range of the estimated post-focus F0 values, i.e., between 86Hz and 90Hz. In the Flat stimuli, the entire contour was set at 90Hz. In the Slightly Declining stimuli, the left edge of the initial vowels were set at 90Hz and the right edge of the final vowels were set at 87Hz, with a straight F0 line inserted between the two edges.

4.2. Procedure: the identification task

Twenty-one English listeners, thirty Japanese listeners, and twenty-seven Seoul Korean listeners participated in the identification task. The English listeners were all one-year or one-semester exchange students studying at Doshisha University. Except for one British student from England, they were all native listeners of American English. The Japanese listeners were mostly from the Kansai area, i.e., the central part of Japan, where Kyoto-Osaka Japanese or dialects similar to it are spoken, and they were all full-time students at Doshisha University. Nineteen of the Seoul Korean listeners were from Ewha Womans University, who had been visiting Doshisha

Univeristy in a two-week study program, and the rest of them were full-time students at Doshisha University Center for Japanese Language and Culture. None of them had reported any hearing disorders. All of them were paid for their participation.

They were all given a questionnaire for training purposes before they participated in the perception task. Within the questionnaire, there was an explanation for the stress alternation of the noun-verb pairs that they were going to listen to in the perception task so that the participants could be aware in advance that stress locations contribute to the noun-verb distinction. The questionnaire also contained the list of sentences in which a pair of words such that one member has a stress mark on the initial syllable and the other member has the mark on the final syllable were embedded, e.g. *The* $\left\{\begin{matrix}\textit{ímport}\\ \textit{impórt}\end{matrix}\right\}$ *of chemicals is banned, Our company* $\left\{\begin{matrix}\textit{ímports}\\ \textit{impórts}\end{matrix}\right\}$ *chemicals from China,* etc. Those embedded words were all those which would be presented to the participants in the following perception task. The participants were, then, asked to choose which member in the pair matches each sentence frame. Through this process, the participants were made accustomed to the notion of stress, that of stress alternation, and the words that they were going to listen to.

The identification task was designed and carried out via SuperLab Version 4.5 installed on a laptop computer (MacBook Air with OS X Version 10.7.4). The three types of stimuli <Natural>, <Flat> and <Slightly Declining> were presented to participants in separate blocks in the order shown in (4). In between those ‘target’ blocks, ‘filler’ blocks with truncated words were inserted. The truncated words in the filler blocks consist of the initial syllable or the first two to three syllables of multisyllabic words. Some of them were created from the stimuli used in the current study, e.g., *TRANS*- from *TRANSplant* and *trans*- from *transPLANT* while others were created from words irrelevant to the current study, e.g. *CAM*- from *CAMpus* and *cam*- from *camPAIGN*, *DOmi*- from *DOminating* and *domi*- from *domiNAtion*, etc. The results obtained from those truncated words in the filler blocks will be considered in my future study, and their details are not laid out here. The order of stimuli presentation within each block was randomized for each participant, and each stimulus was presented only once to each participant. In each block, stimuli presentation was preceded by a short practice session.

(4) Presentation order of blocks

<Natural > → (truncated words) → <Slightly Declining > → (truncated words) → <Flat>

The participants listened to the stimuli through headphones (SONY dynamic stereo headphones MDR-Z500) connected to a computer (MacBook Air, OS X 10.7.4) with a comfortable sound level in a sound-attenuated room. Whenever each stimulus was played to a participant, the letter strings of the word pair to which the stimulus belonged was presented on the computer monitor. In the visual presentation, the trochaic member colored in yellow was placed above the iambic member colored in blue. The location of primary stress was marked with an accent mark ‘′’ above the letter that corresponded to the vowel with primary stress.

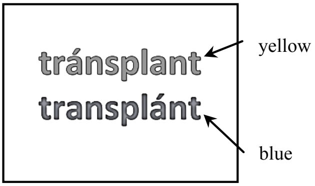

Figure 5. A sample of visual presentation on the computer monitor

Participants were asked to judge which word in the pair they had heard, the trochaic member or the iambic one. They pressed the yellow button of the computer keyboard when they thought that the stimulus they had heard was trochaic, and the blue button when they thought that it was iambic. They were instructed to use only the index finger of their dominant hand when pressing those buttons.

The number of responses in each block per participant was 16 (8 pairs×2 stress patterns). Multiplying the number by blocks and participants, 3,744 (16 responses×3 blocks×78 participants) were obtained in total.

4.3. Analyses

The main goal of the analysis here is to examine whether it is only the Japanese group among the three language groups that show a strong bias towards final stress when presented with the stimuli with the Flat and the Slightly Declining contours. For that purpose, I obtained for each participant (i) the rate of final-stress responses to the stimuli originally produced as iambic (Hit responses) and that of final-stress responses to those originally produced as trochaic (False Alarm (FA) responses), and (ii) the bias measure '*c*'. The bias measure *c* has been proposed in the signal detection theory (Macmillan & Creelman, 1990, 2005) and shows how easily participants respond that the stimuli presented to them have a certain signal. In the current study, the 'certain signal' is iambic (final) stress. To obtain *c* values, the rate (proportion) of Hit and that of FA obtained from each participant were first converted into *z* scores, i.e., the units in the standard normal distribution with a mean of 0 and a standard deviation of 1.[17] A proportion of 0.5 is transformed into a *z* score of 0, and proportions above 0.5 are transformed into positive z scores while those below 0.5 into negative z scores. The z score of Hit and that of FA are denoted as z(Hit) and z(FA) respectively. Secondly, the average of the sum of z(Hit) and z(FA) was obtained, which was further turned into minus, as shown in (5). The average of the sum of z(Hit) and z(FA) indicates the magnitude of the perceiver's bias, and c's distance from zero in the minus range is proportionate to it. In a situation where the perceivers are biased towards neither direction, c reaches near 0.

(5) The formula to obtain c

$$c = -\ \frac{z(\mathrm{Hit})+z(\mathrm{FA})}{2}$$

I calculated c values of each participant in each of the two conditions (Flat and Slightly Declining) separately for the two-syllable words, e.g., *transplant*, and the three-syllable words, e.g., *intercept*.

4.4. Results

4.4.1. The control case: the Natural stimuli condition

Before going into the main results, let us first examine the participants' performance in the Natural stimuli condition, i.e., when enough F0 information was present. Figure 6 summarizes their correct responses in the condition. It shows that the participants' correct response rates far exceeded chance level and reached about 90% for both the trochaic and the iambic stimuli regardless of their native language differences and syllable number differences of the stimuli. That is, participants of all three language groups were good at detecting correct primary stress locations when the F0 information to distinguish the two stress patterns was available.

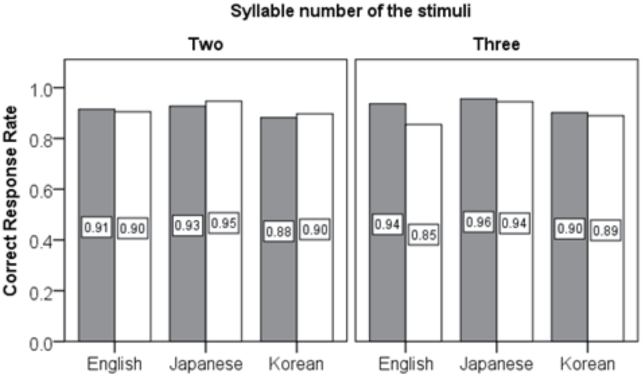

Figure 6. The rate of correct responses in the Natural stimuli condition. The gray bars are of the stimuli produced as trochaic nouns (e.g., *TRANSplant*) and the white bars are of those produced as iambic verbs (e.g., *transPLANT*).

With this in mind, in the next section, let us look at their performance in the Flat and the Slightly Declining condition where the trochaic and the iambic forms were not distinguished by F0.

4.4.2. The Flat and the Slightly Declining stimuli condition

The rates of correct responses to the Flat and the Slightly Slanting stimuli are shown in Figure 7.

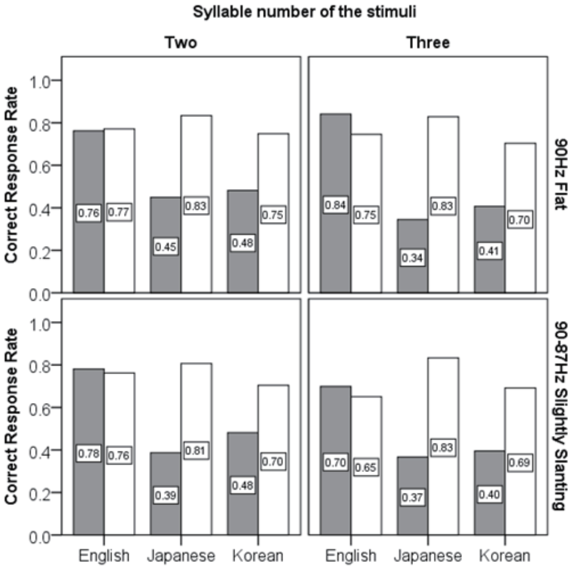

Figure 7. The rate of correct responses in the Flat and the Slightly Slanting stimuli condition. The gray bars are of the stimuli produced as trochaic nouns (*e.g., TRANSplant*) and the white bars are of those produced as iambic verbs (*e.g., transPLANT*).

The native English listeners' correct response rates were as high as about 70% or even more, regardless of whether the stimuli were originally produced as trochaic or iambic. This means that the native English listeners were not biased towards any of the stress patterns, and their *c* values are expected to fall near zero, which will be shown later. The Japanese listeners and the Seoul Korean listeners showed high correct responses rates, as high as about 80% and 70% respectively, to the stimuli originally produced as iambic. When it comes to the stimuli originally produced as trochaic, the Japanese and the Korean listeners' correct response rates declined below 50%. This means that the Japanese and the Korean listeners were biased towards iambic stress, and their *c* values are expected to be in the mins range.

Figure 8 shows the mean *c* values averaged across the participants in each of the three language groups. The mean *c* values of the native English listeners are around zero while those of the Japanese and the Korean listeners are negative in all conditions. This outcome guarantees that the Japanese listeners' and the Korean speakers' responses were biased towards final stress while the native English listener's responses were not.

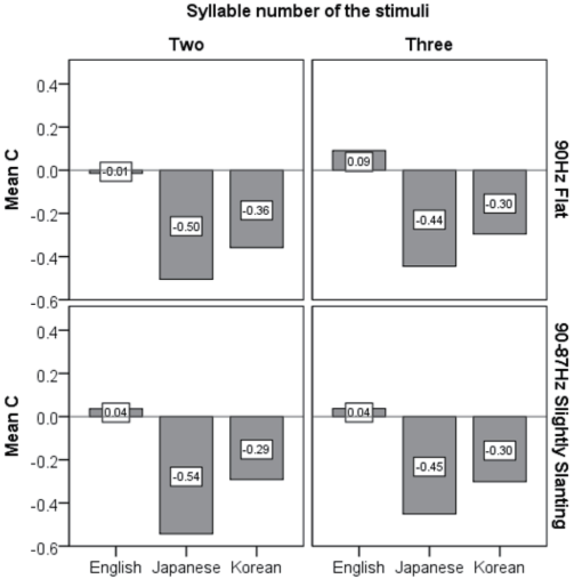

Figure 8. Mean c values

Although the Japanese and the Seoul Korean listeners were similar in that both groups exhibited negative *c* values in all conditions, the absolute values of the Japanese listeners' *c* were consistently greater than those of the Seoul Korean listeners'. This indicates that the Japanese listeners had a stronger bias towards iambic stress than the Korean listeners. This is already evident from Figure 7 that the Japanese listeners' Hit rates, i.e., the rates of correct responses to the stimuli originally produced as iambic, were about 10% higher than the Seoul Korean listeners' (Japanese: about 80%; Korean: about 70%) and the Japanese listeners' Correct Rejection (CR) rates, i.e., the rates of correct responses to the stimuli originally produced as trochaic, were about 5% lower than the Korean listeners' (Japanese: 34 to 45%; Korean: 40 to 48%). That the CR rates of the Japanese group were lower than those of the Korean group means that the Japanese listeners' FA rates were higher than the Korean listeners'.

An ANOVA was carried out using SPSS to test whether the difference in the mean values of *c* between the Japanese and the Korean listeners was statistically significant, in which *c* was the dependent variable and 'language groups', 'F0 contour types', and 'syllable numbers' were the fixed factors. The significance level was set at .05. There was a significant effect of language groups [$F(2, 300) = 30.16$, $p < .001$] while no such effect of F0 contour types and syllable numbers [F0 contour types: $F(1, 300) = 0.002$, $p = .97$; syllable numbers: $F(1, 300) = 0.9$ $p = .34$]. There were no interactions in any combinations of those three factors [language*F0: $F(2, 300) = 0.09$, $p = .92$; language*syllable numbers: $F(2, 300) = 0.08$, $p = .93$; F0*syllable numbers: $F(1,300) = 0.2$, $p = .66$; language*F0*syllable numbers: $F(2, 300) = 0.15$, $p = .86$]. Tukey-Kramer post-hoc comparisons were carried out to see which combination of the languages shows statistically significant difference in the mean *c* values, and every comparison turned out to be significant. Table 8 summarizes the results of the post-hoc tests.

Table 8. The results of Tukey-Kramer comparisons among language groups

Language (I)	Language (J)	Mean difference (I-J)	Standard Error	*p*	95% Confidence Interval	
					lower limit	upper limit
English	Japanese	0.52*	0.067	< .001	0.36	0.68
	Korean	0.35*	0.069	< .001	0.19	0.51
Japanese	English	-0.52*	0.068	< .001	-0.68	-0.36
	Korean	-0.17*	0.063	< .02	-0.32	-0.03
Korean	English	-0.35*	0.069	< .001	-0.51	-0.19
	Japanese	0.17*	0.063	< .02	0.03	0.32

From the fact that the difference between the mean *c* values of the Japanese group and those of the Korean group was significant, it can be concluded that the Japanese listeners were more strongly biased towards final stress than the Korean listeners.

5. Discussion and conclusion

There are three major findings in the results presented above. First, all the three language groups performed well in distinguishing trochaic and iambic stimuli when F0 information to distinguish stress patterns was available. Secondly, the native English listeners were not biased towards initial stress in the Flat and the Slightly Declining condition contra Sugahara's (2011) observation. The third point is that although not only the Japanese listeners but also the Seoul Korean listeners showed a bias towards final stress in the Flat and the Slightly Declining contours, the Japanese group showed a stronger bias than the Korean group.

The first point above guarantees that the participants of all the three language groups were aware of the notion of stress, and that they took the tasks with an intention to earnestly locate the stress positions in the stimuli presented to them. From this, I could safely conclude that the results obtained in this experiment are reliable. It is also necessary to pay attention to the fact that in spite of the lack of lexical stress/accent contrast in Seoul Korean, the Seoul Korean listeners were good at discriminating primary and secondary stress as long as F0 cues to discriminate the two stress levels were available. I will come back to this point later.

As for the second point, one possible account is that the stimuli used in the current experiment contained more noticeable acoustic cues other than F0 to distinguish the trochaic and the iambic patterns than those used in Sugahara (2011), and the native English listeners in the current study might have found it easier to detect stress locations even when no reliable F0 cue to discriminate the two stress patterns was available. For example, the durations of the initial vowels in this study were consistently longer in the primary-stressed syllables than in the secondary-stressed syllables as shown in Table 4. In Sugahara (2011), however, the initial vowels with primary stress and those with secondary stress had almost the same durations in three of the six pairs of the stimuli, and one of the pairs even had an opposite durational relation between the two vowels. Having said that, there is a minor glimpse of an initial stress bias in the results of the native English listeners obtained this time: their correct response rates were 2 to 9% greater for the trochaic stimuli than for the iambic stimuli in three of the four conditions as shown in Figure 6, and the *c* values of those cases were positive as shown in Figure 8, which means that they had some tendency to be biased towards the direction of initial stress though very minor.

The third point is directly related to the question asked in this study. The result that the magnitude of the Japanese listeners' bias towards final stress was significantly greater than that of the Seoul Korean listeners' supports the hypothesis suggested in this study that Japanese listeners' perception of English lexical stress is influenced by a language-specific factor, i.e., the Japanese listeners' knowledge of the antepenultimate accent in their L1. A question is then raised as to what exactly the Japanese listeners' knowledge associated with the antepenultimate accent is and what is the process of inducing their bias. Although I have repeated throughout this article that it is the antepenultimate 'rule' that influences their perception for convenience sake, I do not deny an alternative view that what is actually involved here is the Japanese listeners' stochastic knowledge that the commonest accent location in Japanese is antepenultimate, and they are inclined to assimilate what they hear, even the stimuli of a non-native language, to the forms that they hear most frequently in their L1. Since it is out of the scope of this article to give answer to the debate over the rule vs. the frequency-related knowledge, I leave the question open at this point.

It is further necessary to account for the fact that the Seoul Korean listeners also showed a bias towards iambic stress. One possible account is that the durations of the final syllables were consistently longer than those of the initial syllables in the current stimuli regardless of whether the stimuli was originally produced as trochaic or iambic (see Tables 4 and 5). It might be the case that they paid attention to the relative durational difference between the initial and the final syllables, and responded that the final syllables were more prominent. Another possibility is that they perceived what they heard most recently as the most prominent. It is imaginable that the final syllables remained most salient in the listeners' memory when they were about to make judgments on stress locations. Further investigations are necessary on this point. Another finding about the Seoul Korean listeners is that they also used F0 cues to discriminate the iambic and the trochaic forms just as the Japanese and English listeners did in the Natural-stimuli condition with abundant F0 information. As introduced in Section 3, Seoul Korean also employs F0 in a regular way to mark minor phrase edges and to back up segmental contrast at word-initial positions. With this heavy usage of F0 in their L1, the Seoul Korean listeners might have developed a sensitivity to F0 information in the course of language acquisition, which prevails even in L2 stress perception.

In conclusion, although much remains to be done, the result of the current experiment was in line with the hypothesis that Japanese listeners' perception of English lexical stress is influenced by the antepenultimate accent in Japanese, which is especially unmarked in loanwords. That is, the perception of L2 stress is also partially governed by the distribution of lexical accent in L1.

Acknowledgement

This work was supported by JSPS KAKENHI (Grant-in-Aid for Scientific Research (C)), Grant Number 24520558, which was given to the author. An earlier version of this study without Soul Korean listeners' data were presented at the poster session of ICPP 2013 held at NINJAL in January, 2013, and I thank Shigeto Kawahara for his comment there.

Notes

1 Beckman's (1986) (and Slujiter & Heuven's (1996)) results imply that native English listeners perform well in discriminating different stress patterns in English

even when no or little reliable F0 cue is available, which contradicts the result obtained by Sugahara (2011) that English listeners were not able to properly discriminate English trochaic and iambic words when F0 was not a reliable cue to distinguish the two stress patterns. As will be discussed in Section 5, it is likely that non-F0 acoustic cues for lexical stress such as duration and intensity in Sugahara's (2011) stimuli were, by chance, not salient enough for the native English listeners to pay attention to, and that they were unable to perceive those non-F0 cues there.

2 As Hayes (1995) and Flemming (2007) point out, there are exceptional cases where unstressed syllables carry full vowels. The representatives of such cases are word-final [iː] and [oʊ] as in ˈ*pi.ty and po.*ˈ*ta.to*.

3 As Gussenhoven (1983) briefly summarizes in his introduction, the term 'nuclear accent' has been also referred to as "'nuclear syllable' (Crystal, 1969), 'tonic' (Halliday, 1967), 'sentence stress' (Schmerling, 1976), '[1 stress]' (Chomsky & Halle, 1968), and 'Designated Terminal Element (Liberman & Prince, 1977)" (p.377).

4 Gussenhoven (1984, 1985), however, shows that there are cases where non-final lexical words bear a nuclear accent in broad-focus contexts, and has proposed SAAR (the Sentence Accent Assignment Rule). The rule states that it is the direct argument of a predicate in a focused domain that bears a nuclear pitch accent. Under this rule, even a sentence-initial subject may bear a nuclear pitch accent when the subject is followed by an intransitive (unaccusative) verb such as *disappear* in a broad-focus context.

5 The asterisk '*' means that the tone is linked to a stressed syllable.

6 The sharp pitch drop at the end of the voicing period here is due to its creakiness but not due to tonal configuration. Creaky voice accompanies aperiodic glottal pulses and is intrinsically associated with lower pitch. (Gordon & Ladefoged 2001, Melvyn & Clopper 2015).

7 Some of the Japanese dialects have only one pitch pattern across all lexical items. For example, all lexical items in dialects spoken in Southern Tohoku/Northern Kaonto, e.g., Yamagata, Sendai, Fukushima, Ibaragi, and those spoken in the central part of Kyushu, e.g., Sasebo and Miyazaki, are reported to be uniformly accentless (Kubozono, 2012; Uwano, 2012; among others). When it comes to dialects spoken in the eastern part of Southern Kyushu such as Miyakonojo and Shibushi, the final mora of every prosodic word carries an H tone just as in French (Kubozono, 2012; Uwano, 2012; among others). This could be considered as either a non-contrastive accent on the final mora of every prosodic word or an edge tone at the final position

of every prosodic word. If the latter is the case, then those dialects are also part of the accentless dialects.

8 Although Kyoto-Osaka Japanese is similar to Tokyo Japanese in having both accented and unaccented words, it has an additional parameter, i.e., register tones (*shiki oncho*). Register tones are those which determine the pitch level of an entire prosodic word. There are two types of register tones in Kyoto-Osaka Japanese: high and low. When unaccented, the high register tone spreads till the end of a prosodic word whereas the spreading of the low register tone ends at the end of the penultimate mora and pitch rises at the final mora. When accented, the spreading of register tones is blocked at the accented mora. Other dialects with register tones are those spoken in Shikoku, Noto Peninsula and areas between Kansai and Noto, which are all distributed in the central part of Japan (Uwano, 2012).

9 The dictionary does not show which lexical group a given word belongs to. Therefore, I identified the lexical group of each word one by one. I did not include in my count those with mixed affiliation, i.e., those consisting of morphemes belonging to different lexical groups, such as *mukizu* 'no injury', where *mu* 'none' belongs to the Sino Japanese stratum and *kizu* 'injury' belongs to the Yamato stratum.

10 Here, too, I only extracted accented nouns whose accent patterns are agreed on by all the six speakers in Sugito (1995).

11 Although some of the three-mora loanwords in Kyoto-Osaka Japanese such as *ko.ko˥.a* 'cocoa', *ko.re˥.ra* 'cholera', *go.ri˥.ra* 'gorilla', *to.ma˥.to* 'tomato', *po.te˥.to* 'potato', *ra.ji˥.o* 'radio', *ro.si˥.a* 'russia' have an accent on their penultimate mora, which are considered to represent the most Kyoto-Osaka-like accent pattern, they are in fact limited in number. See Tanaka (2009) for more details.

12 Although Korean had been originally a language with lexical accents or tones, accentual or tonal contrasts had been lost in most dialects by the end of the 16th century (Kim, 2013). There are a few dialects in modern days that still keep lexical accents, e.g., South and North Kyungsang (Kenstowicz & Sohn, 2001; Jun, 2006; Lee, 2009; Kang, 2010; Davis, Tsujimura & Tu, 2012; Kim, 2013), which are spoken in the eastern part of the Korean peninsula. Kim (2013), however, states that these dialects are also losing tonal contrasts nowadays and becoming more similar to Seoul Korean.

13 Jun uses the term 'accentual phrase', instead of the term 'minor phonological phrase'. The two terms have been used interchangeably, especially in the studies of Japanese prosody. In this section on Seoul Korean, I choose to use the term 'minor

phonological phrase' because the term 'accentual phrase' may evoke a wrong idea that Seoul Korean has a lexical accent.

14 Although minor phonological phrase formation in Korean may vary due to various factors such as the presence or absence of narrow focus, it usually consists of a noun+particle and of a main verb by itself in a broad focus condition (Kim, 2013).

15 Korean obstruent consonants have three-way contrasts: aspirated, fortis (tense) and lenis (lax): "Lenis stops, /p t k/, in general have weak release bursts, moderately long VOT (35±90 ms) with weak aspiration energy, and breathy vowel onset after the consonant; aspirated stops, /p^h t^h k^h/, have strong release bursts and long VOT (75±120 ms) with strong aspiration energy; tense stops, /p' t' k'/, have strong release bursts and short VOT (0±35 ms)" (Jun, 1998, p.205).

16 To obtain spectral balance, I used a long-term average spectrum of each vowel interval with a bandwidth of about 300Hz though Plag et al (2011) used a bandwidth of 100Hz.

17 When the rate is 1 or 0, z scores cannot be computed. In those cases, the rate was converted by using the following calculus proposed by Macmillan & Creelman (2005): z(1) was converted to z(1-0.5/n) and z(0) was converted to z(0.5/n), where n stands for the number of the relevant iambic-trochaic pairs (see also Kawahara & Shinohara, 2015).

References

Akinaga, K. (Ed.). (1981). *Meikai akusento jiten* [Meikai accent dictionary] (2nd ed.). Tokyo: Sanseido.

Beckman, M. E. (1986). *Stress and non-stress accent*. Dordrecht: Foris Publications.

Beckman, M. E., & Edwards, J. (1994). Articulatory evidence for differentiating stress categories. In P. A. Keating (Ed.), *Phonological structure and phonetic form: papers in Laboratory Phonology III* (pp. 7-33). Cambridge: Cambridge University Press.

Bolinger, D. L. (1958). A theory of pitch accent in English. *Word*, *14*, 109–149.

Carr, P. (2012). *English phonetics and phonology*. Hoboken, NJ: Wiley.

Chomsky, N., & Halle, M. (1968). *The sound pattern of English*. New York: Harper & Row.

Coltheart, M. (1981). The MRC psycholinguistic database. *Quarterly Journal of*

Experimental Psychology, *33A*, 497-505.

Cooper, N., Cutler, A., & Wales, R. (2002). Constraints of lexical stress on lexical access in English: evidence from native and non-native listeners. *Language and Speech*, *45*, 207-228.

Crystal, D. (1969). *Prosodic systems and intonation in English*. Cambridge: Cambridge University Press.

Cutler, A., & Carter, D. M. (1987). The predominance of strong initial syllables in the English vocabulary. *Computer Speech & Language*, *2*, 133-142.

Dainora, A. (2006). Modeling intonation in English: a probabilistic approach to phonological competence. In L. Goldstein, D. H. Whalen & C. T. Best (Eds.), *Laboratory Phonology 8* (pp. 107-132). Berlin, New York: Mouton de Gruyter.

Davis, S., Tsujimura, N., & Tu, J-Y. (2012). Toward a taxonomy of loanword prosody. *Catalan Journal of Linguistics*, *11*. 13-39.

de Jong, K. (2004). Stress, lexical focus, and segmental focus in English: patterns of variation in vowel duration. *Journal of Phonetics*, *32*(4), 493-516.

Dupoux, E., Pallier. C., Sebastian, N., & Mehler, J. (1997). A destressing “deafness” in French? *Journal of Memory and Language*, *36*(3), 406-421.

Flemming, E., & Johnson, S. (2007). Rosa's roses: reduced vowels in American English. *Journal of the International Phonetic Association*, *37*, 83-96.

Fujisaki, H., Hirose, K., & Sugito, M. (1986). Comparison of acoustic features of word accent in English and Japanese. *Journal of the Acoustical Society of Japan*, *(E)7*, 57–63.

Gordon, M., & Ladefoged, P. (2001). Phonation types: across-linguistic overview. *Journal of Phonetics*, *29*. 383-406.

Gussenhoven, C. (1983). Focus, mode and the nucleus. *Journal of Linguistics*, *19*, 377-417.

Gussenhoven, C. (1984). *On the grammar and semantics of sentence accents*. Dordrecht: Foris.

Gussenhoven, C. (1985). Two views of accent: a reply. *Journal of Linguistics*, *21*, 125-138.

Gussenhoven, C. (2004). *The phonology of tone and intonation*. Cambridge: Cambridge University Press.

Gussenhoven, C., & Bruce, G. (1999). Word prosody and intonation. In van der Hulst, H. (Ed.), *Word prosodic systems in the languages of Europe* (pp. 233-271). Berlin: Mouton de Gruyter.

Halliday, M.A.K. (1967). *Intonation and grammar in British English*. The Hague: Mouton.

Hayes, B. (1995). *Metrical stress theory*. Chicago: The University of Chicago Press.

Hyman, L. (2006). Word-prosodic typology. *Phonology*, *23*, 225-257.

Jun, S-A. (1998). The Accentual Phrase in the Korean prosodic hierarchy. *Phonology*, *15*, 189-226.

Jun, S-A. (2005). Korean intonational phonology and prosodic transcription. In S-A. Jun (Ed.), *Prosodic typology* (pp. 201-229). Oxford: Oxford University Press.

Jun, S-A. (2006). Intonational phonology of Seoul Korean revisited. In T. Vance & K. Jones (Eds.), *Japanese Korean Linguistics 14* (pp.15-26). Stanford: CSLI.

Kang, Y. (2010). Tutorial overview: suprasegmental adaptation in loanwords. *Lingua*, *120*, 2295-2310.

Kawahara, S. (2015). The phonology of Japanese accent. In H. Kubozono (Ed.), *The handbook of Japanese phonetics and phonology* (pp. 445-492). Berlin: De Gruyter Mouton.

Kawahara, S., & Shinohara, K. (2015). Iconic inferences about personality: from sounds and shapes. In M. K. Hiraga, W. J. Herlofsky, K. Shinohara & K. Akita (Eds.), *Iconicity: East meets West* (pp.57-70). Amsterdam: John Benjamins Publishing Company.

Kenstowicz, M., & Sohn, H. (2001). Accentual adaptation in North Kyungsang Korean. In M. Kenstowicz (Ed.), *Ken Hale: a life in language* (pp. 239-270). Cambridge: MIT Press.

Kim, K. (2013). *Tone, pitch accent and intonation of Korean: a synchronic and diachronic view* (Doctoral dissertation). Universitat zu Koln.

Kubozono, H. (2001). Epenthetic vowels and accent in Japanese: facts and paradoxes. In J. van de Weijer & T. Nishihara (Eds.), *Issues in Japanese phonology and morphology* (pp. 113–142). Berlin, New York: Mouton de Gruyter.

Kubozono, H. (2006). Where does loanword prosody come from? A case study of Japanese loanword accent. *Lingua, 116*, 1140-1170.

Kubozono, H. (2012). Varieties of pitch accent systems in Japanese. *Lingua*, *122*. 1395-1414.

Ladd, D. R. (1980). *The structure of intonational meaning: evidence from English*. Bloomington: Indiana University Press.

Ladd, D. R. (2008). *Intonational phonology* (2nd ed.). Cambridge: Cambridge University Press.

Lee, D. (2009). *The loanword tonology of South Kyungsang Korean* (Doctoral dissertation). Indiana University.

Liberman, M., & Prince, A. (1977). On stress and linguistic rhythm. *Linguistic Inquiry*, *8*, 249-336.

Macmillan, N. A., & Creelman, C. D. (1990). Response bias: characteristics of detection theory, threshold theory and "nonparametric" indexes. *Psyhological Bulletin*, *107*(3), 401-413.

Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: a user's guide* (2nd ed.). Mahwah: Lawrence Erlbaum Associates Publishers.

McCawley, J. (1968). *The phonological component of a grammar of Japanese*. The Hague: Mouton.

Melvin, S., & Clopper, C. G. (2015). Gender variation in creaky voice and fundamental frequency. *Proceedings of ICPhS 2015*.

Okobi, A. (2006) *Acoustic correlates of word stress in American English* (Doctoral dissertation). MIT.

Peperkamp, S., & Dupoux. E. (2002). A typological study of stress 'deafness'. In C. Gussenhoven & N. Warner (eds.), *Laboratory phonology 7* (pp.203-240). Berlin: Mouton de Gruyter.

Pierrehumbert, J. (1979). The perception of fundamental frequency declination. *Journal of the Acoustical Society of America*, *66*, 363-369.

Pierrehumbert, J. (1980). *The phonology and phonetics of English intonation* (Doctoral dissertation). MIT.

Pierrehumbert, J., & Beckman, M. (1988). *Japanese tone structure.* Cambridge, MA: MIT Press.

Plag, I., Kunter, G., & Schramm, M. (2011). Acoustic correlates of primary and secondary stress in North American English. *Journal of Phonetics*, *29*, 362-374.

Schmerling, S. F. (1976). *Aspects of English sentence stress*. Austin: Texas University Press.

Selkirk, E. (1980). The role of prosodic categories in English word stress. *Linguistic Inquiry*, *11*, 563-605.

Selkirk, E. (1984). *Phonology and syntax*. Cambridge, MA: MIT Press.

Selkirk. E. (1995). The interaction of constraints on prosodic phrasing. In M. Horne (Ed.), *Prosody: theory and experiment* (pp. 231-262). Dordrecht: Kluwer Academic Publishers.

Shattuck-Hufnagel, S. (1995). The importance of phonological transcription in empirical approaches to "stress shift" versus "early accent": comments on Grabe and Warren, and Vogel, Bunnell and Hoskins. In B. Connell & A. Arvaniti (Eds.), *Phonology and phonetic evidence: papers in Laboratory Phonology IV* (pp 128-140). Cambridge: Cambridge University Press.

Shibata, T. (1994). Gairaigo-ni okeru akusentokaku-no iti [The position of accent nucleus in loanwords]. In K. Sato (Ed.), *Gendaigo/hougen-no kenkyuu* [The study of the modern Japanese language and dialects] (pp. 388-418). Tokyo: Meiji-Shoin.

Sluijter, A. M. C., Shattuck-Hufnagel, S., Stevens, K. N., & van Heuven, V. J. (1995). Supralaryngeal resonance and glottal pulse shape as correlates of stress and accent in English. *Proceedings of ICPhS 2* (pp. 630-633).

Sluijter, A. M. C., & van Heuven, V. J. (1996). Acoustic correlates of linguistic stress and accent in Dutch and American English. *Proceedings of ICSLP*, 2.

Sugahara, M. (2011). Identification of English primary stress and bias toward strong word-initial syllables: native vs. Japanese listeners. *Proceedings of ICPhS 2011* (pp. 1918-1921).

Sugahara, M. (2012). Phonetic evidence for prosodic word prominence in American English. In T. Borowsky, S. Kawahara, T. Shinya & M. Sugahara (Eds.), *Prosody matters: essays in honor of Elisabeth Selkirk* (pp. 208-234). Sheffield: Equinox.

Sugito, M. (1995). *Osaka/Tokyo akusento onsei jiten* [Pronunciation dictionary of Osaka/Tokyo accent]. Tokyo: Maruzen.

Tanaka, S-I. (2009). Osaka hogen gairaigo-no akusento-to shiki-ni tuite [On accent and *shiki* in Osaka Japanese loanwords]. *Proceedings of the 138th meeting of Linguistic Society of Japan* (pp.214-219).

Terken, J., & Hirschberg, J. (1994). Deaccentuation and persistence of grammatical function and surface position. *Language and Speech*, *37*(2), 125–145.

Van Leyden, K., & van Heuven, V. J. (1996). Lexical stress and spoken word recognition: Dutch versus English. In M. den Dikken & C. Cremers (Eds.), *Linguistics in the Netherlands* (pp. 159-170). Amsterdam: John Benjamins.

Uwano, Z. (2012). Three types of accent kernels in Japanese. *Lingua*, *122*, 1415-1440.