Universiteit Utrecht

Master Thesis

Cross-linguistic Perception of Pitch Position

Shuangshuang Hu

4104676

Linguistics: the Study of the Language Faculty

UiL-OTS

August 2015

Supervisor: Prof. dr. R. W. J. Kager, UiL-OTS

Second Reader: dr. A. Chen, UiL-OTS

Abstract

Pitch contrasts signal word meanings by pitch contours (rise versus fall), pitch levels (high versus low) or positions (earlier or later in a word). Previous studies have largely focused on native or non-native perception of pitch contours and pitch levels (Gandour 1983; Wang et al. 1976; Hallé et al. 2004; Xu et al. 2006; among others). Few studies have been directed towards perception of pitch positions. To fill in the gap, the current study made an attempt to investigate perception of pitch position by listeners of different word prosodic systems, Dutch listeners (stress system) and Japanese listeners (pitch accent system). Dutch resembles Japanese in that they both use positional marking to signal lexical items, but they differ by the means: Dutch exploits the location of stress while Japanese employs pitch accent for positional marking. The two languages also differ in the role of pitch function at the word level. Pitch is one of the acoustic correlates in Dutch stress while pitch in Japanese is exploited exclusively to distinguish lexical meanings. Given these, the current study aimed to examine whether Dutch listeners whose native languages do not have lexically contrastive pitch were able to perceive non-native pitch position contrasts at the acoustic level and phonological level. An ABX discrimination task and a sequence recall task were implemented in the study for the purpose of examining listeners' acoustic perception and phonological perception, respectively. Results of the two experiments were discussed with respect to the predictions from different perceptual models. The study made an attempt to provide an understanding of how pitch processing is influenced by native languages.

Content

Chapter 1. Introduction

Pitch in human language is notoriously multi-functional: it conveys our attitudes, emotions, or knowledge states, and it also carries word meaning in many languages. Multi-functionality of pitch poses a big problem for listeners, but on top of this, languages also show much variation in how meaningful pitch contrasts are signaled phonetically: by pitch levels (high versus low), pitch contours (rise versus fall), or positions (earlier or later in a word). Multi-functionality and variability entail that humans have a remarkable ability to *tune in* to different functions of linguistic pitch amidst an ocean of phonetic variation, suggesting that our speech perception systems are *selectively sensitive* to pitch contrasts, depending on our native language, as well as, possibly, intrinsic language-independent properties of pitch perception. In order to reveal how our perceptual system copes with pitch contrasts in word prosody, a considerable number of studies have been directed towards native or non-native perception of pitch levels (Gandour 1978; Guion 2004; Wayland & Guion 2004; Schaefer & Darcy 2014; among others) or pitch contours (Wang 1976; Hallé et al. 2004; Xu et al. 2006; among others) by listeners from different language backgrounds. However, compared with perceptual studies on pitch levels and pitch contours, fewer have focused on pitch position.

In order to fill in the gap, the current study is focused on investigating how listeners of Dutch and Tokyo Japanese, featuring two different prosodic systems, process pitch position. Tokyo Japanese only uses a contrastive position of pitch fall (a “pitch accent”, from high to low) in word prosody for lexical contrasts, for example *haNA* meaning “flower” and *HAna* meaning “nose” (upper case here denotes a pitch accented syllable), which is used as a baseline to examine pitch position perception in the current study. Dutch exploits the position of stress to signal a lexical contrast, for example, *VOORnaam* (upper case hereinafter denotes a stressed syllable) meaning “first name” and *voorNAAM* meaning “respectable”, which, in a sense, resembles Japanese in that both the languages use positional marking to signal lexical items. However, Dutch and Japanese differ in the way they use linguistic pitch for positional marking: Dutch uses stress and Japanese uses a pitch fall at the lexical level. On the other hand, the two languages differ in how linguistic pitch functions in word prosody: pitch is used exclusively in Japanese to signal contrasts in lexicon while it is one of the acoustic correlates of stress in Dutch. Given these, the current study attempts to disentangle whether Dutch listeners are able to process non-native pitch position contrasts at the acoustic and phonological levels, and how the native prosodic system influences pitch position perception. To achieve this goal, two experiments were implemented that aimed to examine listeners' perception at the acoustic and phonological processing levels, respectively. The study attempts to provide an understanding of how pitch processing is influenced by native language in listeners with different word prosodic systems. This may make a humble contribution to building a comprehensive view of the dynamic function of pitch in human languages and how phonological knowledge is used in processing spoken language.

The current study contains six chapters. Chapter 2 offers a general review of perceptual studies on pitch contour, pitch level in the perspective of native and non-native perception on the three pitch dimension. Chapter 2 also compares Dutch and Japanese word prosody with respect to positional marking and the functional use of lexical pitch. In order to make testable predictions, Chapter 2 discusses perceptual models and the predictions they make. Chapter 3 discusses the specific research questions addressed in the current study. Chapters 4 and 5 describe two experiments (an ABX discrimination task and a sequence recall task) designed to investigate listeners' perception at the acoustic and phonological levels, respectively. Chapter 6 provides a discussion and conclusions of the study and points out the limitations in the study as well.

Chapter 2. Literature review

Pitch is used universally for intonation, distinguishing, for example, between question and statement. At the same time, however, pitch functions very differently across languages in how it is used to signal word meaning. Languages can be classified as having one of four prosodic systems according to the role pitch plays in signaling lexical contrasts. A language can have a lexical tonal system (e.g., Mandarin, Thai), a pitch-accent system[1] (e.g., Japanese), a lexical stress system[2] (e.g., English, Dutch, and Spanish) or a prosodic system without lexically-contrastive pitch (e.g., Korean, French). This chapter is organized into three sections. The first section contains a general review on how listeners of different word prosodic systems discussed above cope with native and non-native pitch contrasts in three dimensions, pitch levels, pitch contours and pitch positions. It discusses a) perception of pitch level and pitch contour, first with respect to native perception and then non-native perception, and b) perception of pitch positions. The second section introduces characteristics of Dutch and Japanese word prosody. The third section discusses perceptual models and predictions regarding the research questions in the current study.

2.1. Perception of pitch contrasts

2.1.1 Perception of pitch level & pitch contour

Pitch level and pitch contour are two distinctive dimensions of lexical tone. They play a crucial role in native and non-native perception of lexical tones in listeners of different language backgrounds (Burham & Francis 1997; Brham et al. 1992; Lee et al. 1996; Vũ 1981; Gandour 1983; Lee & Nusbaum 1993; Wang et al. 1999; Wang et al. 2006; among others). This section first reviews native perception of the two dimensions by listeners whose native language features pitch contour or/and pitch level in word prosody. Non-native perception of pitch level and pitch contour is reviewed after native perception so that it may shed light on how listeners exploit phonological knowledge from their native language to cope with non-native pitch contrasts.

[1] Here the notion “pitch-accent” has been applied to languages as widely different as Swedish (word stress, contrastive tone only on stressed syllable) and Japanese (no stress, no contrastive tone levels nor contours, but only a contrastive position of a pitch fall). Pitch in “a pitch-accent system” is the primary correlate of prominence and there are significant constraints on the pitch patterns for words (Bybee et al. 1998). Note that Hyman (2009) argued there is no phonological prototype of “pitch accent” but a restricted tonal system.

[2] Note that lexical stress can mean two things: (a) stress at the word level or (b) stress that is lexically contrastive. In the current study, both properties are relevant, hence a lexical stress system will be defined as one in which stress is lexically contrastive at the word level.

2.1.1.1 Native perception[3]

Contrasts in pitch level and/or pitch contour are exploited to minimally distinguish lexical items in tone languages, which makes tone languages advantageous for studies regarding pitch perception (Gandour 1978; Francis et al. 2008; Repp et al. 1990). In tone languages such as Mandarin and Thai, word meaning changes as pitch contour/level changes within the domain of the monosyllable. For instance, tones in Mandarin are traditionally cited as Tone 1, 2, 3, 4 where pitch pattern varies in terms of high-level, rising, dipping and falling contour, respectively (Howie 1976). The meaning of a monosyllable /ma/ changes with respect to pitch patterns Tone 1, Tone 2, Tone 3, and Tone 4 as "mother", "numb", "horse" and "to blame", respectively. Different from Mandarin, which features one level tone (Tone 1) and three contour tones (Tone 2,3,4) differing in pitch contour, Thai features three level tones differing in pitch level (high, mid, low) and two contour tones differing in pitch contour (rising, falling) (Abramson 1979).

Empirical studies on perception of pitch level and/or pitch contour have focused on a) whether native tone listeners perceive native tone contrasts categorically, and b) whether having native lexical tone helps them perceive tone contrasts in various acoustic/linguistic environment (e.g., acoustic variability, limited acoustic input, contextual condition) (Hallé et al. 2004; Gottfried & Suiter 1997; among others). When perceiving contrasts of native pitch contours, Mandarin listeners were found to perceive the pitch contour contrasts in a categorical manner (Wang 1976; Hallé et al. 2004; Xu et al. 2006). To be more specific, a pitch continuum ranging from a level tone to a contour tone (Tone 1 vs. Tone 2) was perceived categorically by Mandarin listeners (Wang 1976). Hallé et al (2004) made a further investigation on perceiving different pairs of contrasts (Tone 1 vs. Tone 2, Tone 2 vs. Tone 4, and Tone 3 vs. Tone 4 in Mandarin) by Mandarin Taiwanese listeners, via a classic identification task (AX) and a discrimination task (ABX). They found quasi-categorical perception in Mandarin Taiwanese listeners of the tone contrasts regardless of a continuum ranging from level tone to contour tone or contour tone to contour tone. Different from Mandarin listeners, perception of Thai tones by native Thai listeners, given the differences in pitch level (high, mid, low), was not categorical (Abramson 1979).

Studies on the one hand have revealed that having native lexical tone can facilitate perception of native tone contrasts when acoustic information is varied or missing (Blicher et al. 1990; Gottfried & Suiter 1997; Lee 2009; among others). For instance, Lee (2009) found that Mandarin listeners succeeded in identifying four isolated tones with acoustic variability (in voices of 16 male speakers and 16 female speakers). He also found they were able to track and identify isolated contour differences even when the acoustic information of the tone was

[3] Empirical studies reviewed in this section discuss native listeners of lexical tonal system (e.g., Mandarin, Thai) perceiving native pitch dimensions in their native languages.

missing (i.e., only the onset of a monosyllable). Likewise, Mandarin listeners were found to achieve remarkable accuracy in perceiving incomplete (i.e., only onset) tone contrasts (Tone 1 vs. Tone 4, Tone 2 vs. Tone 3) (Lee et al. 2008). It is proposed that Mandarin listeners may have shaped categorical boundaries for each tone, which helps them to perceive the pitch contour contrasts regardless of missing acoustic information or acoustic variability (Chan et al. 1975; Lee 2009; Xu et al. 2006). However, interestingly, Thai listeners had trouble in perceiving contrasts of pitch level (mid vs. low) in isolated Thai tones with acoustic variability (in voices of 5 male speakers and 5 female speakers) (Abramson 1979). The finding implies that Thai listeners might not only rely on pitch level, but also on pitch register in voice range to discriminate native Thai tones (Abramson 1976; Schaefer & Darcy 2013). On the other hand, having native lexical tone does not necessarily facilitate native pitch perception. When the pitch patterns of two pitch levels or pitch contours are similar to each other in certain respects, it yields difficulty for native tone listeners (Leather 1990; Burnham et al 1996; among others). Mandarin listeners were found to have difficulty in perceiving modified fragmented tone contrasts that share similar onset pitch height (Tone 1 vs. Tone 4, both having high onset pitch; Tone 2 vs. Tone 3, both having low onset pitch) (Lee et al. 2008). Thai listeners, in an AX discrimination task, showed the most difficulty in discriminating the tone contrasts between pitch levels (mid vs. low) whereas discrimination between pitch contours (rising vs. falling) was the easiest (Burnham et al. 2000, Schaefer & Darcy 2013). This is because mid tone and low tone in Thai, unlike high, rising and falling tones, both "are typically characterized by very little movement of fundamental frequency over time" (Gandour 1978: 43).

To conclude, on the one hand, native listeners of different tone languages differ in perception of their native pitch contrasts. Mandarin listeners perceive pitch contours in Mandarin tones categorically while Thai listeners do not show categorical manner in perceiving native pitch levels in Thai tones. Also, Mandarin listeners are able to identify tone contrasts regardless of pitch register in voice range or missing/variable acoustic input while Thai listeners rely on pitch register as well as pitch level to distinguish tone differences, which implies that perception of native pitch contrasts may be influenced by long-term categorical/non-categorical representation in native languages (Abramson 1979, Xu et al. 2006). On the other hand, native listeners of tone languages show confusion in perceiving native pitch contrasts when the pitch patterns of pitch levels and/or pitch contours are similar to each other in certain respects (e.g. similarly high/low onset, very few fundamental frequency movement).

2.1.1.2 Non-native perception

Previous studies have revealed that non-native perception is influenced by phonological knowledge in native languages (Wayland & Guion 2004; Schaefer & Darcy 2013, 2014;

among others). In order to discuss non-native perception on pitch contour and/or pitch level, this section, according to the use of lexically contrastive pitch in the native language, divides listeners of different prosodic systems into two groups, a) listeners whose native language has lexically contrastive pitch such as lexical tone (e.g., Mandarin) and pitch accent (e.g., Tokyo Japanese), and b) listeners whose native languages do not have lexically contrastive pitch such as stress[4] (e.g. Dutch, Spanish, English) and neither stress nor any other lexically contrastive pitch (e.g., French, Seoul Korean).

Non-native perception by listeners whose native languages have lexically contrastive pitch

Having lexically contrastive pitch in one's native language may facilitate perception of non-native pitch contrasts (Gandour 1983; Wayland & Guion 2004; Schaefer & Darcy 2013, 2014; among others). For instance, Wayland & Guion (2004) found that Mandarin listeners significantly outperformed English listeners in perceiving contrasts of non-native pitch levels (low vs. mid) in Thai in discrimination tasks in short (500 ms) and long (1500 ms) interstimuli interval (ISI) conditions both before and after auditory training. And interestingly, Mandarin listeners improved significantly in discrimination at the longer ISI condition after training while English listeners did not (Wayland & Guion 2004). This suggests that the ability to track pitch direction, movement or change at word level in the native language may be transferable to the perception of non-native tones (Gandour 1983; Xu et al. 2006). Among the very few studies that have investigated non-native pitch perception by Japanese (pitch-accent system) listeners, Schaefer & Darcy (2013, 2014) found that Japanese listeners showed a better performance than stress listeners (English listeners) in perceiving contrasts of pitch level and pitch contour in Thai tones. They proposed that the functional use of lexically contrastive pitch helps Japanese listeners to track pitch level/contour contrasts in Thai.

On the other hand, however, having lexically contrastive pitch in the native language does not necessarily help in non-native perception of pitch contrasts (Wayland & Guion 2004; So 2006; So & Best 2010). For example, Mandarin listeners had trouble in perceiving modified Thai tone contrasts of pitch level (mid vs. low) that differ slightly in pitch patterns (Wayland & Guion 2004). This failure may not be caused by a lack of exposure to Thai tones, but by a lack of a one-to-one mapping between the two Thai tones (mid tone vs. low tone) and the closest Mandarin tones. So (2006) examined perception of pitch contour contrasts in Mandarin tone by Cantonese (Mandarin naïve) listeners via an identification task, and found Cantonese listeners had difficulty in perceiving Tone 1 vs. Tone 4 in that they misidentified Tone 4 as Tone 1, which indicated that they may have mapped both Mandarin Tone 1 (high

[4] Pitch naturally plays some role in stress system, but it is not lexically contrastive (Hayes 1995; Sluijter & van Heuven 1996; Kager 1989; Trommelen & Zonneveld 1999; among others).

level) and Tone 4 (falling) into one tone category Cantonese tone 1 (high falling). The failure of Cantonese listeners in discriminating pitch contour contrasts in Mandarin may be due to assimilation of two non-native contrastive tones into one tonal category in native language (So & Best 2010).

In addition to the influence of native language on non-native perception, pitch level and pitch contour themselves weigh differently for native tone listeners when perceiving non-native tones. Contrasts of pitch level (high vs. mid vs. low) in Thai were more difficult than those of pitch contours (rising vs. falling) for Mandarin listeners (Schaefer & Darcy 2014). Likewise, Gandour (1983) compared native tone listeners of Thai, Mandarin, Cantonese and Taiwanese on perception of 19 different synthesized non-native tones recorded in one voice (5 level tones, 14 contour tones where beginning and end points varied). Listeners, after listening to the stimuli twice, were required to differentiate pairs on an 11-point scale (from no difference to extreme difference). The findings showed that all the listeners used pitch level significantly more than pitch contour to judge tone dissimilarity though pitch contour was also used to a large extent. Gandour’s finding (1983) was in accordance with that of Tuc (2003) which argued that in perceiving non-native modified tones contrastively in both pitch level and pitch contour, pitch level was a more salient cue for listeners to discriminate pitch contrasts.

In sum, listeners whose native languages have lexically contrastive pitch are able to use their ability in native language to track non-native pitch level or pitch contour and transfer their native pitch contrasts in word prosody onto non-native tones (Wayland & Guion 2004; Xu et al. 2006). The influence of phonological knowledge in native language does not necessarily facilitate them to perceive non-native pitch contrasts in the condition that they lack a one-to-one mapping between native tone and non-native tone (So 2006; So & Best 2010).

Non-native perception by listeners whose native languages do not have lexically contrastive pitch

Languages that do not have lexically contrastive pitch include a) lexical stress languages in which pitch is one of the acoustic correlates of stress but is not marked contrastively[5] in lexicon (e.g., Dutch, Spanish, and English), and b) languages that have neither lexical stress nor other lexically contrastive pitch, e.g., French and Seoul Korean. Previous studies have demonstrated that listeners of lexical stress systems such as English focused on acoustic properties of pitch differences in tonal contrasts when processing non-native pitch contour

---

[5] Dutch stress is not lexically contrastive (Kager 1989; Hayes 1995; Trommelen & Zonneveld 1999; among others). According to Cutler (1986), minimal stress pairs are very rare. She identified 13 clear minimal stress minimal pairs based on the CELEX lexical database (Baayen et al. 1993).

and pitch level (Wayland & Guion 2004; Wu & Lin 2008; So & Best 2010; among others). Wu & Lin (2008) used an AX identification task to examine English listeners' perception of Mandarin tone contrasts and found that they tended to show psycho-acoustic listening. In other words, they paid close attention to the similarities in pitch offset and onset between pitch level contrasts (Chan et al. 1975). Likewise, Canadian English listeners had greater problems in identifying tone contrasts in Mandarin Tone 1 vs. Tone2, Tone 2 vs. Tone 3 and Tone 1 vs. Tone 4 than other contrasts, which may be due to the similarities between these contrasts in onset pitch height or in pitch patterns (So & Best 2010).

Moreover, listeners whose native languages do not feature lexically contrastive pitch are not "deaf" to non-native pitch contour or pitch level, but their perception is not categorical. For instance, English listeners were found to be more sensitive to small fundamental frequency variations, which are irrelevant to categorizing Mandarin tones, while Mandarin listeners ignored those subtle and irrelevant tonal variations to categorize pitch contours into tones efficiently (Stagray & Downs 1993). Dutch listeners were also observed to perceive Mandarin tones contrasting in pitch contour in a psycho-acoustic manner, unlike Mandarin listeners whose perception is categorical (Chen & Kager 2012). Listeners whose native language features neither stress nor lexically contrastive pitch such as French listeners were found sensitive to non-native pitch contour differences in Mandarin, but they failed to perceive pitch contour contrasts in a set of linguistic categories; instead, their perception was also psycho-acoustic, like stress listeners (Hallé et al. 2004).

The discussed findings of non-native perception of pitch contour and/or pitch level are subject to two concerns. One is that listeners whose native languages do not have lexically contrastive pitch may perceive pitch contrasts in non-native tones as melodic contours, which do not have any phonemic status in their native languages (Hallé et al. 2004; So & Best 2010). The other concern is that listeners may map the pitch contour/level contrasts onto intonational patterns, for example, mapping a rising pitch contour in Mandarin onto question intonation patterns or mapping a falling pitch contour onto statement intonation patterns in their native languages (Francis et al. 2008).

In conclusion, listeners of different prosodic systems behave differently in perceiving native or non-native contrasts of pitch level and/or pitch contour. In native perception by tone language listeners, Mandarin listeners perceive Mandarin pitch contour contrasts in a categorical manner while Thai listeners do not have categorical perception in perceiving Thai pitch level contrasts. In non-native perception, listeners whose native languages do not have lexically contrastive pitch such as Dutch listeners and French listeners are not "deaf" to tone differences but they perceived tone contrasts in a psycho-acoustic manner. Listeners' perceptual ability depends on their native languages. Having contrastive pitch at word level enables native tone listeners to use their ability depending on their native languages to track pitch variations such as pitch movement, pitch change or pitch height in non-native pitch perception (Guion 2004; Wayland & Guion 2004; Xu et al. 2006). It is also proposed that non-native perception is shaped by the functional use of lexical pitch in their native languages (Wang et al. 2006; Schaefer & Darcy 2013, 2014).

2.1.2 Perception of pitch position

Compared with perceptual studies on pitch contour and pitch level, no previous studies have investigated the perception of pitch position. The studies that come closest to this are Dupoux's studies on non-native perception of lexical stress. Dupoux et al. (1997 2001) applied a sequence recall task with memory load and phonetic variables (e.g., different voices) to investigate the perception of stress location in Spanish by French listeners. They controlled the stressed syllable with a longer duration (by 7 ms longer), a higher pitch (by 96 Hz) and louder decibel (by 6.1 dB) compared with unstressed syllable in minimal pairs. Perception of stress location can be regarded as perception of position in essence. They found that French listeners had great difficulty discriminating non-words that differed only in the location of stress marking. In French, stress does not carry lexical information, but predictably falls on the word's final vowel. French listeners do not need to process stress to identify lexical items (Rietveld 1980). The finding in Dupoux et al. (1997 2001) revealed that French listeners who don't have lexical stress in their native language were deaf to stress position contrasts, though they are not tone-deaf (Hallé et al. 2004).

This raises an interesting question: are listeners whose native language does not have lexically contrastive pitch able to process positional marking of pitch? In order to answer the question, the current study investigates whether listeners of a stress language, Dutch, with positional marking in stress, are able to perceive contrasts of pitch positions in Tokyo Japanese, a language only using the position of pitch accent marking (no stress, no contrastive tone contours or tone levels) to signal lexical contrasts.

2.2 Dutch & Japanese word prosody

Dutch has word stress and stress in Dutch is not lexically contrastive (Kager 1989; Hayes 1995; Trommelen & Zonneveld 1999; van Oostendorp 2012; among others). Minimal stress pairs such as in (1)[6] differing only in the position of stress, not in vowel quality or segment content, are rare in Dutch, according to Cutler (1986), who found only 13 clear minimal stress pairs were identified from CELEX (Baayen et al. 1993).

(1) kaNON "gun" [ka'nɔn]

[6] The examples are from van Oostendorp, M. 2012: 344.

CAnon "norm" ['kanɔn]

Studies have revealed that Dutch listeners exploit stress location for word recognition in minimal pairs (Cutler & van Donselaar 2001; Sluijter & van Heuven 1996; among others). For instance, Cutler & van Donselaar (2001) found that Dutch listeners used stress to constrain lexical access in minimal pairs. In their study, *VOORnaam* ("first name") and *voorNAAM* ("respectable") only activated the meaning of *VOORnaam* and *voorNAAM*, respectively, instead of activating both meanings of the two words.

A considerable number of studies have focused on acoustic realization and perceptual features of stress to signal lexical items in Dutch in unaccented or accented condition (van Heuven 1988; Sluijter et al. 1995; Cutler & Donselaar 2001; Sluijter & Heuven 1995, 1996; Quené & Koster 1998; among others). Based on cross-linguistic studies, vowel duration, intensity and pitch have been shown to be acoustic correlates of stress. A stressed syllable can be distinguished by a combination of increases in loudness and duration as well as fuller phonetic quality, indicated by an "absence of spectral reduction" (Sluijter & van Heuven 1996:2471). Dutch speakers were found to employ vowel duration as the most salient cue to signal lexical contrasts in disyllabic real words and nonwords that only differ in stress position (Sluijter & van Heuven 1996). Based on a linear discriminate analysis (Nooteboom 1972; Eefting 1991), Sluijter & van Heuven (1995, 1996) examined Dutch native listeners' discrimination between lexical stress on initial position vs. final position in two conditions: accented and unaccented. They found that vowel quality and overall intensity were the poorest cues of stress position while duration was the most effective correlate, unaffected by accent. Loudness was a reliable cue even in the unaccented condition, which was close in strength to duration.

Other studies have focused on which cues of stress influence word recognition in Dutch. Dutch listeners appear to use vowel duration as the most reliable cue to segment words (Quené 1992; Quené & Koster 1998). In a gating task (Grosjean 1980) conducted by van Heuven (1986), Dutch listeners showed a perceptual bias in perceiving stress, in that they tended to assume the first syllable was stressed even when the actual stressed syllable occurred in later positions. Additionally, stress on the medial position had the most influence on Dutch listeners when segmenting words. Stress on the first syllable had the least effect on perception (Sluijter & van Heuven 1996).

Tokyo Japanese resembles Dutch in that both the languages use positional marking in word prosody. However, different from Dutch which uses stress for positional marking, Japanese uses pitch accent, an abrupt pitch fall from high to low (HL pitch fall), to contrast word meaning. Pitch accent is used in two ways to signal lexical contrasts: first, whether or not it is present, and second, if present, where it is located (Kawahara 2015). In the examples below in (2)[7], the pitch accent is marked with an apostrophe (') on the syllable[8] it occurs. For instance, the meanings of /ha ʃi/ in (2a) ("chopsticks") and (2b) ("bridge") differ in the location of the pitch accent while "ha ʃi" in (2c) ("edge") is differentiated from (2a) and (2b) in absence vs. presence of the pitch accent.

(2). Accentual patterns: on the first syllable   on the second syllable   no pitch accent

Tonal patterns: H'L   LH'   LH

a.	b.	c.
ha ʃi o nu ru	ha ʃi o nu ru	ha ʃi o nu ru
"chopsticks-accu paint"	"bridge-accu. paint"	"edge-acc.paint"
("paint the chopsticks"	"paint the bridge"	"paint the edge")

The underlying pitch accentual patterns are mapped onto surface tonal patterns. Japanese uses only two levels of tonal heights lexically: high tone (H) and low tone (L). According to the Initial Lowering Rule (Haraguchi 1977), generally speaking, a multisyllabic[9] word in Tokyo Japanese begins with a L tone, and the second and consecutive unaccented syllables[10] have H tone, creating a LHHH pattern. If the first syllable is accented, it bears a H tone and all the other syllables following it receive L tone. Phonetically speaking, when pitch accent falls on a syllable (not the first syllable) in a multisyllabic word, the accented vowel receives a H tone, and the syllable preceding (not the first syllable) is also assigned a H tone while the following syllable is assigned a L tone. It is notable that an unaccented word and the word with accent on the final position in 2(b) and 2(c), only differing in the absence of pitch fall are minimal pair. Indeed, the two words in the minimal pair are phonetically similar when they appear in isolation (Vance 1995; Warner 1997). The final tone in an unaccented word is slightly lower

[7] The example is extracted from Hayes. 2008: 292.

[8] The current study focuses on perception of pitch position in the domain of regular mora (syllable). A mora is a subsyllabic unit. For instance, *shinshinto* (meaning "silent") has three syllables (*shin, shin, to*) but five morae (*shi, n, shi, n, to*). However, many morae are indeed also syllables: the allowable structures such as V, CV, CCV (Otake & Cutler 1993; Kubozono 1989). Heavy syllables (a regular mora (a syllable) + a special mora (i.e., nasal coda, doubled consonant)) are not discussed here.

[9] [10] Indeed, mora is the tone-bearing unit in Japanese. The notion "multisyllabic" and "syllable" here refer to "multimoraic" and "mora", respectively. The current study focuses on perception of pitch position in the domain of regular morae type: CV (i.e., syllables) (Otake & Cutler 1993; Kubozono 1989).

(but still higher than any of the other tones in the word) than that in the word with final syllable accented. The slight tonal differences on the final position in a minimal pair can be realized by attaching a nominative particle suffix [+ga], as illustrated in (3).

(3). a. /kaki’ + ga/ (final accented) “fence + NOM”

```
          |  |        |
          L H’    L
```

b. /kaki +ga/ (unaccented) “permission + NOM”

```
       |  |      |
       L  H    H
```

According to Haraguchi (1999), if there are *n* (the number) syllables in a word, then there are (*n* + 1) possible accentual patterns in Tokyo Japanese. Taking a four-syllable word as an example, each pattern is provided as followed[11]:

(4).

Surface tonal patterns	Position of pitch accent	examples	meaning
(a) H’LLL	Initial syllable	ka’makiri	mantis
(b) LH’LL	Second syllable	hima’wari	sunflower
(c) LHH’L	Third syllalbe	tamane’gi	onion
(d) LHHH’	Final syllable	imooto’	sister
(e) LHHH	Unaccented	gakusee	student

Contrasts of pitch positions are illustrated in example (4) in accordance with tonal patterns. For instance, the H tone accented in 4(a) contrastively differs that in the position in 4(b); the H’L tonal combination in 4(a) differs in the position of that in 4(b) and 4(c). 4(d) and 4(e) differ in the final position in which the H tone in 4(d) is accented while that in 4 (e) is unaccented. Tonal differences contrastive in positions are determined by the position and/or absence of pitch accent. The positional contrasts caused only by lexically contrastive pitch in Tokyo Japanese make it a good baseline to test perception of pitch positions in the current study.

To conclude, Dutch and Japanese are similar in that they both make use of positional marking to signal word meaning, but by different means. Dutch uses positional marking of stress in word prosody, while Japanese uses pitch-accent for positional marking. Given the similarities and differences, it may be informative to ask whether Dutch listeners are able to perceive contrasts in tonal positions in Tokyo Japanese. In order to answer this question, perceptual models for making testable predictions are discussed in the following section.

2.3 Theoretical perceptual models

Current theoretical models with respect to non-native perception have been proposed based on abstract phoneme (Best 1994, 1995; Best et al. 1988, 2001; Best & Tyler 2007),

[11] The examples used are from Kawahara. S. 2015: 457.

phonological features (Brown 1998, 2000) or the function of phonetic or phonological dimension in native language (McAllister et al. 2002). These models[12] have largely focused on non-native perception of segmentals such as vowels and consonants and may be applicable as well to the perception of suprasegmentals (i.e., tone, pitch accent, stress and intonation).

2.3.1 Perceptual Assimilation Model

Listeners are confronted with discrepancies between the properties of the non-native sounds and those of native phonemes when presented with a speech contrast that is not employed by the native language. The Perceptual Assimilation Model (PAM) posits that listeners assimilate articulatory similarities of non-native phonemes to their native categories, which results in different types of assimilation (Best 1994, 1995; Best et al. 1988, 2001; Best & Tyler 2007). Assimilation types depend on the comparison between the listeners' native phonemic inventory and the non-native sound. To be more specific, various assimilation patterns depend on the divergence of non-native phones[13] from native categories. Listeners perceive a non-native sound as whether it can be categorized into their native categories, which means the performance in discrimination differs depending on how well the assimilated non-native phones are fit to the native category (Best 1995; Best et al. 1988, 2001). The assimilation types relevant for this study are mainly Single Category (SC) assimilation and Two Category (TC) assimilation. SC assimilation occurs when the two non-native phones are mapped equally well onto the same single category in listeners' native inventory, in which cases PAM predicts poor discrimination in that the two non-native phones are perceived as equally good examples of the same native category. TC assimilation, on the other hand, occurs when the two non-native phones are mapped onto two native categories that are contrastive. In the cases of TC assimilation, PAM predicts good discrimination in that the two non-native phones are mapped onto one native contrastive pair, and the two non-native phones in such a pair are phonemically distinct from each other, thus resulting in easier discrimination.

PAM successfully predicted non-native perception of segmentals (Goto 1971; Miyawaki et al. 1975; Sheldon & Strange 1982; Best et al. 1998; Rohena-Madrazo 2013; among others) and has been applied for non-native perception of suprasegmentals as well (So 2006; So & Best 2010). For instance, Japanese listeners were found to have difficulty in perceiving the contrast of /l/ - /ɹ/ in English (Sheldon & Strange 1982). Japanese has only one liquid while

---

[12] Note that Speech Learning Model (SLM) (Flege 1995) is not discussed here in that SLM was originally developed to predict and explain perception of segments in second language acquisition. It posits non-native sounds are related to articulatory positions ("position sensitive allophones": Flege 1995: 238), rather than to a more abstract phonemic level (also see Rohena-Madrazo 2013 for comparisons between SLM and PAM).

[13] The notion "phone" here refers to an individual sound unit of speech without concern as to whether it is a phoneme of some language.

English has two. The non-native /l/ and /ɹ/ may be assimilated onto SC in Japanese native inventory, which resulted in poor discrimination. Rohena-Mardrazo (2013) found that Spanish listeners were able to discriminate contrasts of non-native occluded intervocalic voiced stops /VtV/-/VdV/ (V stands for "vowel") in Brazilian Portuguese in that non-native occluded voiced stops were perceived as a pair of TC assimilation. Furthermore, predictions of PAM have been examined in the scope of non-native perception of suprasegmentals (So 2006; So & Best 2010). For example, So & Best (2010) examined whether PAM can be applicable to cross-linguistic perception of non-native tone contrasts in Mandarin by Mandarin-naïve Cantonese listeners, Japanese listeners and Canadian English listeners. They applied an identification task on six pairs of tone contrasts in Mandarin and found that Cantonese listeners had difficulties in perceiving Tone 1 vs. Tone 4 but they were good at discriminating Tone 1 vs. Tone 2. Their poor discrimination of Tone 1 vs. Tone 4 may be because Cantonese listeners assimilated Mandarin Tone 1 (high level tone) and Tone 4 (falling tone) onto the Cantonese Tone 1 (high level tone) as a Single Category. Tone 1 in Cantonese features two allotones, high level and high falling in pitch contour. When two non-native phones (Mandarin Tone 1 and Tone 4) assimilate equally well or poorly to a single native phonemic category, SC occurs and results in poor discrimination. As for the good discrimination of Tone 1 vs. Tone 2 pair, Cantonese listeners may perceive this contrast as a TC assimilation pair. Mandarin Tone 1 and Tone 2 may be assimilated to the Cantonese Tone 1 and Tone 2, which resulted in good discrimination.

In addition to the successful predictions in non-native perception discussed above, PAM faces challenges due to considering suprasegmental dimensions as categories in the same way as segmental categories (Hao 2008; So & Best 2010; Schaefer & Darcy 2013; Dupoux et al. 2001, 2008). In particular, PAM faces unsolved issues in two respects: a) assimilation from one non-native category of suprasegmental dimension (e.g., tone) onto another native suprasegmental dimension (e.g., pitch-accent, stress), and b) mapping one non-native category of suprasegmental dimension (e.g., tone) onto an element of the native intonation system. In So & Best (2010), PAM failed to predict non-native tone perception by listeners of Japanese (featuring pitch-accent) and English (featuring stress). In their study, Japanese listeners were found to have difficulties in perceiving isolated Mandarin Tone 2 and Tone 4 pairs. Tone 2 as rising pitch contour and Tone 4 as falling pitch contour are similar to pitch-accent patterns LH' and H'L in Japanese bisyllabic words, respectively. PAM predicts that Tone 2 and Tone 4 may be perceived as a TC assimilation pair, which should have resulted in good discrimination. However, the finding in So & Best (2010) contradicted the predictions PAM made. Besides, in the same study of So & Best (2010), English listeners were observed to have difficulties identifying Tone 1 vs. Tone2, Tone 2 vs. Tone 3 and Tone 1 vs. Tone 4. It is argued that non-native tone assimilation for stress listeners is complicated in that they may perceive tone contrasts as nonlinguistic melodic variations (Hallé et al. 2004) or they may perceive tones as “certain kinds of prosodic categories” (So & Best 2010: 289). English listeners may perceive non-native tones as nonlinguistic melodic changes similar to the ones they use in prosodic patterns such as intonation patterns, but note that these prosodic patterns do not have “any linguistic significance” like those of lexical tones or pitch accents and they do not have “phonemic status” in English (So & Best 2010: 289). Alternatively, English listeners may map the two non-native tones such as Tone 2 and Tone 4 onto stress patterns Strong-Weak and Weak-Strong, respectively, which is in accordance with the finding that English listeners were able to discriminate Tone 2 vs. Tone 4 contrast (So & Best 2010). However, it is yet unclear of the assimilation between other pitch contour contrasts in Mandarin to prosodic categories in English and mapping between intonation and tones.

Another issue with respect to PAM is *dissimilation* between perception of non-native suprasegmentals and native category (Dupoux et al. 1997, 2001; Schaefer & Darcy 2013, 2014). In the framework of PAM, if one non-native suprasegmental contrast does not exist in native inventory, it is to be perceived dissimilar to the native phoneme and is not assimilated onto any native phonological system, which results in good discrimination (So & Best 2010, Dupoux et al. 2008). However, this prediction fails to support the finding in Dupoux et al. (2008) that French listeners were “deaf” to stress in Spanish. French does not have lexical stress, and according to PAM, non-native stress will not be assimilated to any existing French phonological category, which should have resulted in easier discrimination.

In sum, PAM successfully predicts non-native perception of segmentals and can be applied to non-native perception of suprasegmentals. However, it also faces challenges of conceiving suprasegmental dimensions as categories and mapping between one non-native suprasegmental dimension (e.g., tone) to another native suprasegmental dimension (e.g., pitch accent, stress). Regarding perception of pitch positions in the framework of PAM, it can be predicted that Dutch listeners may assimilate certain contrasts of pitch positions such as H’L vs. LH’ into two categories (e.g., Strong-Weak, Weak-Strong stress patterns, respectively) and thus yield good discrimination. Minimal pairs that contrastive in the final position (i.e., final accented syllable vs. unaccented syllable in the final position), which does not exist in Dutch phonological system, may not be assimilated to any existing categories, which may result in good discrimination. Alternatively, the two non-native phonemes in the minimal pair contrastive in the final position may be assimilated to the same single category (e.g., Weak-Strong stress pattern), because they both have similar surface tonal patterns, which may yield poor discrimination.

2.3.2 Models based on phonological representation

Brown (1998, 2000) developed a model based on phonological interference that explains how influence of the native phonology arises and identifies the level of phonological knowledge that impinges upon non-native perception/acquisition. She postulated that if the learners' native grammar lacks the phonological feature that differentiates a particular non-native contrast, they would be unable to perceive the contrast and therefore unable to acquire the novel segmental/supreasegmental representations.

Brown's model has successfully predicted non-native perception/acquisition of segmentals under the influence of native linguistic experience (Schwartz 1993; Yamada 1996; Brown 2000; Johnson & Babel 2007; among others). For example, Brown (2000) studied the discrimination of English /l/ and /r/ by Chinese and Japanese listeners through an AX discrimination task and a picture selection task (Brown & Matthew 1993, 1998). The results in her study indicated native grammar in Japanese listeners constrained their discrimination of the non-native contrast due to a lack of phonological contrasts in the native grammar, whereas Chinese listeners made successful discrimination of the contrast. Johnson & Babel (2007) examined non-native perception of English intervocalic fricatives /s/, /ʃ/ and /θ/ by Dutch listeners and found Dutch listeners had trouble in discriminating /s/ vs. /θ/ and /ʃ/ vs. /θ/ in that they tended to perceive /s/ and /θ/, /ʃ/ and /θ/ as similar pairs. The difficulty in discriminating these contrasts in Dutch listeners were attributed to the lack of /θ/ in Dutch native inventory (Johnson & Babel 2007). These findings demonstrate that a listener's native grammar may actually *impede* perception/acquisition of a non-native phonological feature, preventing the non-native listener from perceiving/acquiring a non-native phonemic contrast.

Predictions from Brown's model (1998, 2000) have also supported non-native perception of suprasegmentals (Dupoux et al. 1997, 2001, 2008). Dupoux et al. (1997, 2001) investigated French listeners' perception on lexical stress contrasts in Spanish. They found that French listeners were "deaf" to lexical stress, which is successfully predicted by Brown's model (1998, 2000) in that the lack of suprasegmentals (e.g., lexical tone, pitch-accent, stress) at the lexical level in French brings difficulty for French listeners and even French learners of Spanish.

Under the account of the model based on phonological representations, it can be predicted that Dutch listeners may be able to perceive contrasts of pitch positions due to the usage of positional marking of stress in their native language. Dutch uses stress location in the sense that stressed syllables are always marked stronger than unstressed syllables, which shapes an opposition between the stressed and unstressed syllables. However, it can also be predicted that the lack of unmarked position (final unaccented position) in Dutch word prosody may cause difficulty for Dutch listeners in discriminating Japanese minimal pairs, contrastive in the presence and absence of pitch accent marking.

2.3.3 Feature Hypothesis

The Feature Hypothesis, proposed by McAllister et al. (2002) predicts that the more prominently a certain phonetic or phonological dimension functions in the native language, the easier it might be to learn to discern and use that dimension for non-native phonological processing. The hypothesis is supported by the study of non-native perception of lexically-contrastive vowel length in Swedish (McAllister et al. 2002). McAllister et al. (2002) conducted a mispronunciation detection task to examine the perception of non-native isolated Swedish words contrastive in vowel length by listeners of English (in which vowel length exists as a correlate of lexical stress), Estonian (in which has lexically-contrastive vowel length) and Spanish (which does not use vowel length lexically). They found perceptual performance of the three groups showed in a hierarchy: Estonian (performed the best) >> English >> Spanish (performed the worst). They suggested the performance in discriminating non-native vowel length contrasts depended on the use of lexically contrastive phonetic dimension (i.e., vowel length) in native languages.

In addition to the successful predictions on non-native perception of segmentals, Feature Hypothesis is supported by studies on non-native perception of suprasegmentals (Schaefer & Darcy 2013, 2014). Schaefer & Darcy (2013, 2014) examined the perception of non-native isolated Thai tones by listeners of Mandarin (where lexically contrastive pitch functions in monosyllable domain), Tokyo Japanese (where lexically contrastive pitch functions in multisyllable domain), English (where no lexically contrastive pitch but partly lexically contrastive stress) and Seoul Korean (where no lexically contrastive pitch/stress functions). They found a hierarchy in performance, from high to low accuracy rate, among the listeners: Mandarin >> Japanese >> English >> Seoul Korean, which implied that perception of non-native pitch contour/level contrasts in Thai was determined by the functional use of lexically contrastive pitch in native language (Schaefer & Dary 2013). Schaefer & Darcy (2014) proposed that to a more extent the pitch is exploited at the word prosody in native language, the better the listeners will perform in perceiving non-native tones.

Under the account of the Feature Hypothesis, it can be predicted that Dutch listeners may encounter more difficulties than Japanese listeners in perceiving pitch positional contrasts. Pitch functions more prominent in Japanese than in Dutch, according to the Feature Hypothesis, because pitch is one of the acoustic correlates of stress but it does not function to signal lexical contrasts in Dutch, while Japanese exploits lexically contrastive pitch and pitch (i.e., a pitch fall) is exclusively used to signal lexical items.

2.4 Summary

Theoretical perceptual models discussed above mostly account for non-native segmental perception. Models based on abstract phoneme categories such as PAM (Best 1994, 1995; Best et al. 1988, 2001; Best & Tyler 2007) posit that two non-native phones are assimilated to the same single or two existing native category, depending on the discrepancy from native phonemes. PAM successfully predicts some findings on perception of suprasegmentals as well, such as perception of non-native lexical tones by native tone listeners (So & Best 2010).

However, PAM is challenged by some issues regarding assimilation from one non-native suprasegmental dimension category (e.g., lexical tone) to another native suprasegmental dimension (e.g., pitch accent, stress) and mapping non-native suprasegmental categories (e.g., lexical tone) onto intonation patterns (So & Best 2010; Hao 2008). It also faces dissimilation issues (Dupoux et al. 2001, 2008).

On the other hand, models based on phonological representations (Brown 1998, 2000) and the Feature hypothesis (McAllister et al. 2002; Schaefer & Darcy 2013, 2014) have focused on how phonological representation or the function of a phonetic/phonological feature in the native language influences non-native perception/acquisition and have made successful predictions regarding non-native perception on segmentals. These models are also supported by perceptual studies on non-native perception of suprasegmentals though these studies are not as many as those examining predictions from PAM.

Chapter 3. Research questions

As discussed above, word meaning can be signaled by pitch contrasts such as by pitch level (high versus low), pitch contour (rise versus fall), or pitch position (earlier or later in a word). Compared with the considerable number of studies focusing on perception of pitch level and pitch contour (Kiriloff 1969; Bluhme & Burr 1971; Burham & Francis 1997; Brham et al. 1992; Lee et al. 1996; Vũ 1981; Gandour 1983; Lee & Nusbaum 1993; Wang et al. 1999; Wang et al. 2006; among others), fewer studies have been directed towards perception of pitch position. To fill in the gap, the current study is concerned with whether listeners of Dutch, a language in which the position of a stress accent in a word can be lexically contrastive, are able to perceive pitch position contrasts occurring in different tonal patterns in Tokyo Japanese, a language in which the position of a pitch fall in a word is lexically contrastive. Dutch resembles Japanese in that they both use positional marking for lexical contrasts, but differ in the means (stress vs. pitch fall). Indeed, the difference between which suprasegmental dimension they use for positional marking reflects the difference in how linguistic pitch functions at lexical level: lexical pitch is used exclusively in Japanese to signal lexical contrasts while it is one of the acoustic correlates of stress in Dutch.

Therefore, the current study raises the following questions: Do native listeners of a language with a lexically contrastive stress-accent system perceive lexically contrastive pitch positions? In other words, will Dutch listeners show deafness to contrastive pitch positions from Tokyo Japanese? Alternatively, will the positional use of stress marking in Dutch equip them to perceive pitch positions?

Chapter 4. Experiment 1 (ABX discrimination task)

The first experiment aimed to investigate the perception of pitch positions by native listeners of Dutch and Japanese. It examined the two groups of listeners with an ABX discrimination task involving contrasts of tonal positions marked by pitch accent. Subjects were instructed to listen to three tokens that differed only in the pitch accent position and were required to press a button on a button box to indicate whether the third token corresponded to the first or to the second token. In this experiment, contrasts of pitch positions on two-, three and four-syllable items were tested.

4.1 Method

4.1.1 Subjects

43 Dutch native speakers who were naïve to any language that uses lexical tone or pitch-accent and 15 Japanese monolinguals participated in the experiment, with Dutch subjects as the experimental group and the Japanese subjects serving as a control group. Subjects in the Japanese group were from all walks of life including 2 undergraduate students. The Japanese group consisted of 2 females and 10 males, aged from 21 to 30, with an average age of 26, and they all come from the following areas: Saitama, Kanagawa, Tokyo and Chiba, where Tokyo Japanese (standard Japanese) is spoken (Hirayama 1970). None of them had an educational background related to linguistics. The experimental group consisted of undergraduate students from Utrecht University, 33 females and 10 males, aged from 19 to 25, with an average age of 21. Two of the subjects in the experimental group had linguistic background while the others were recruited from non-linguistic/language majors.

4.1.2 Stimuli

Stimuli were nonwords in both languages consisting of two syllables (CVCV), three syllables (CVCVCV) and four syllables (CVCVCVCV). In order to make reliable tokens, properties of vowels and consonants in phonetic repertoire in Dutch and Japanese were taken into account. Dutch has 12 vowels and Japanese has 10 vowels (including 5 long vowels). The two languages differ in consonant inventory and consonant clusters as well. To cope with the differences, only segments legal in both languages were considered. First, five vowels legal in both languages were chosen: /a/, /i/, /u/, /e/ and /o/. Next, it is known that /a/, /i/, /u/ were eliminated in that vowels /i/ and /u/ often have higher f0 and /a/ has lower f0 than the others in natural utterances across languages (Lehiste & Peterson 1961; Ladefoged & Maddieson 1984; Whalen & Levitt 1995). Then regarding acoustic properties of consonants, it is known that high pitch is attracted to the vowel after a voiceless consonant (Lehiste & Peterson 1961; Lehiste 1970). Vowels after aspirated and fortis plosives are marked as H tone and those after lenis explosives as L tone (Kim & Duanmu 1999). Accordingly, /m/, /n/ and /j/

were chosen as consonants for constructing nonword tokens in the current study. Given the concerns discussed above, segments of /ne/, /me/, /no/, /mo/, /je/ and /jo/ were considered as candidates for constructing nonwords. However, /me/, /ne/ and /momo/ have meanings in Japanese, meaning “eyes”, ‘yes’ and “peach”, respectively in Japanese. Therefore, segments /no/ and /jo/ were used for constructing two-, three- and four-syllable nonwords as “nono” and “yoyo”, “nonono” and “yoyoyo”, “nononono” and “yoyoyoyo”. Note that heavy syllables (containing long vowels or geminated consonants) were eliminated to avoid confounding. This is because a heavy syllable accented in word final position is easier to identify than unaccented or accented syllables in non-final positions (Ishizawa 2011). Also, five-syllable words were not considered in the materials due to the fact that most five-syllable real words in Japanese are compound (Shibata 1994; Kubozono 2008; Labrune 2012).

All tokens were constructed in possible phonological tonal patterns in Tokyo Japanese. As mentioned before, tonal patterns are realized by the location and the presence of the pitch accent (HL pitch fall). Based on the Initial Lowering Rule (Haraguchi 1977), the first two syllables in a word carry a L-H tonal sequence, unless the accent falls on the first syllable (phonetically speaking, the first vowel). When the first syllable is accented, a H tone is assigned and the second syllable receives a L tone. An unaccented word and a word with the final syllable accented are regarded as a minimal pair. The minimal pair has an identical tonal pattern. The difference in the minimal pair is that the final syllable in the unaccented word receives a H tone slightly lower than the H tone in final accented word but still much higher than all the L tones in the other syllables in the word (Haraguchi 1977, 1999). Accordingly, tokens carrying all possible tonal patterns of two-, three- and four-syllable words were constructed, as shown in Table 1.The syllable where the accent falls is marked with an apostrophe (‘), while a word is left unmarked when no accent occurs. Two-, three- and four-syllable nonwords result in 3, 6 and 10 pairs of tonal position contrasts, respectively, resulting in 19 pairs of accent contrasts in total, as shown in Table 2.

Pitch accent position	2 syllable	3 syllable	Four syllable
Unaccented	LH	LHH	LHHH
On the final syllable	LH’	LHH’	LHHH’
On the first syllable	H’L	H’LL	H’LLL
On the second syllable		LH’L	LH’LL
On the third syllable			LHH’L

Table 1. All possible tonal patterns in two-, three- and four-syllable condtions.

Contrast in two-syllable nonwords		
LH’ - LH	LH’ – H’L	LH – H’L

Contrasts in three-syllable nonwords		
LHH' - LHH	LHH' - H'LL	LHH' – LH'L
LHH – H'LL	LHH – LH'L	H'LL – LH'L
Contrasts in four-syllable in four-syllable nonwords		
LHHH' - LHHH	LHHH' - HLLL	LHHH' – LH'LL
LHHH' – LHH'L	LHHH – H'LLL	LHHH – LH'LL
LHHH – LHH'L	H'LLL – LH'LL	H'LLL – LHH'L
LH'LL – LHH'L		

Table 2. Contrasts of pitch positions.

All tokens were embedded in a carrier sentence in declarative intonation:

kinoo	______ (targeted nonword)	ga	kita.
"Yesterday	_____ (targeted nonword)	nominative marker	came."

The nominative marker "ga" serves to differentiate the unaccented words and words with final syllable accented (minimal pair) (Vance 1995, Warner 1997). For instance,

(3). a. /kaki' + ga/ (final accented) "fence + NOM"

```
     |  |     |
     L H'   L
```

b. /kaki +ga/ (unaccented) "permission + NOM"

```
     |  |     |
     L  H    H
```

"ga" in 3(a) carries a L tone when the preceding word is accented on the final position, whereas it is assigned a H tone in 3(b) when the preceding word is unaccented. The L tone of "ga" allows the tone in the final syllable in an unaccented word to be assigned a slightly lower tone (though still marked as H) than the H tone in the final syllable accented word.

All tokens were clearly articulated and recorded by a trained female phonetician and checked by another phonetician. Both phoneticians are native speakers of Tokyo Japanese. The targeted nonwords were abstracted from the sentences and were manipulated. Each token was digitized at 16 KHz at 16 bits. Average durations of two-, three- and four-syllable tokens were 96.1 ms, 142.2 ms and 191.8 ms, respectively. Any tokens containing lengthened vowels and/or glottalized final syllables were eliminated.

Each trial in the experiment consisted of three stimuli: A, B and X. X had two identities, either the same as A or as B. Token A and B always had the same segments but differed in tonal patterns marked by pitch accent positions. Contrasts of all possible tonal patterns in each syllable condition were examined. For a given contrast, A and B received two different pitch accent locations, for instance, "nono" in H'L and LH' tonal patterns, or in LH', H'L. This resulted in two different A-B orders (A-B, or B-A). X, the token to be discriminated, had the same segments as A and B, but had the same tonal pattern as A on half of the trials and as B on the other half. The overall stimuli design had the following structure: Token Set ("nono" vs. "yoyo") x Accent Contrast (see in Table 2) x Accent Order (A-B vs. B-A) x X identity (X=A vs. X=B): 2 x 19 x 2 x 2, resulting in 152 trials. The stimuli are attached in the Appendix.

4.1.3 Procedure

All the subjects were instructed to listen to a foreign language. They were told that they would hear three consecutive tokens, A, B, X. The first token A and the second token B were different, but the third token X was identical to either A or B. They were seated in front of a computer, with a button box with button A and button B, representing the first token A and second token B, respectively. After listening to the third token X, subjects were asked to press button A or button B to judge whether the token X corresponded to A or B, respectively. Each trial was separated by an interval of 800 ms. No time pressure was applied in each trial in the experiment but the subjects were required to press the button as fast as possible. The trials of different syllable conditions and pitch contrasts were counterbalanced. The trial proceeded only after the subject finished the previous trial.

The experiment consisted of two phases: practice phase and test phase. The practice phase contained six trials, each syllable condition (two-, three-, four- syllable) involved in two trials. In the six practice trials, subjects received a feedback of "correct" or "incorrect" to indicate whether their response was correct or not after they pressed the button. The feedback came up in the computer screen 500 ms. later after their response and displayed for 1000 ms and then disappeared from the screen. The test phase consisted of one block. No feedback was displayed in the test phase. Responses and reaction times of each subject were measured from the onset of the X stimuli by ZEP software package (Veenker 2012).

4.2 Results

Accuracy rate and reaction time (RT) of each tonal contrast in two-, three- and four-syllable conditions in each subject were analyzed and were transformed to z-scores as well. Based on Van Selst and Jolicoeur (1994), any datum with a z-score beyond the range of (-2.246, 2.246) was regarded as an outlier. In mean accuracy rate, data of two subjects (z-score: -2.53, -2.68, respectively) in Dutch group and of one subject (z-score: -2.73) in Japanese group were eliminated. In mean RT, data of three subjects (z-score: 3.02, 2.98, 3.92) in Dutch group were eliminated.

Firstly, whether Dutch and Japanese subjects showed difference in perceiving tonal contrasts in two-, three- and four-syllable conditions was investigated. Mean accuracy rates and mean RTs in each syllable condition in Dutch and Japanese subjects were analyzed by an

Independent T test and are shown in Tables 3 – 5, respectively. Significant level was set at 0.016 (0.05/three syllable conditions) by Bonferroni correction. In accuracy analysis, Dutch subjects performed as well as Japanese subjects. Both groups achieved high accuracy rates in all the three syllable conditions. No significant mean difference was found between Dutch subjects and Japanese subjects in each syllable condition ($F(1, 56) = 0.95$, $p=0.34$ in two-syllable condition, $F(1, 54) =0.49$, $p=0.56$ in three-syllable condition, $F(1, 54) = 0.81$, $p= 0.77$ in three-syllable condition). Reaction time analysis showed Dutch subjects were significantly slower than Japanese subjects in all the three syllable conditions (two-syllable condition: $F(1, 53) = 12.09$, $p=0.001$; three-syllable condition: $F(1, 53) = 13.87$, $p=0.001$; four-syllable condition: $F(1, 55) = 7.07$, $p=0.004$).

Secondly, whether the two language groups performed differently in each tonal contrast in each syllable condition was examined. Mean accuracy rates of each tonal contrast in each syllable condition were analyzed by an ANOVA, respectively. The two-syllable condition contains three tonal contrasts, which were subjected to three within-subject dependent variables. Similarly, six tonal contrasts in three-syllable condition, ten tonal contrasts in four-syllable condition were subjected to six and ten, respectively, within-subject factors. Language background was between-subject factor in each syllable condition. Table 6 and 7 report that both Dutch subjects and Japanese subjects performed quite equally in discriminating tonal contrasts in two-syllable and three-syllable conditions. No significant mean difference was observed in all the tonal contrasts in the two- and three-syllable conditions. Table 8 reveals that Dutch subjects made significantly more errors than Japanese subjects in perceiving LHHH vs. LHHH' contrast in four-syllable condition (Accuracy rate: 63.60% vs 81.93%, $F(1, 1) = 21.40$, $p <0.001$). No significant mean difference was observed between the two language groups in perceiving other tonal contrasts in four-syllable condition. In addition, a T test showed that the performance of the two groups in each contrast in each syllable condition was significantly beyond chance level.

Thirdly, whether syllable length (two-, three-, four-syllable) affected perceptual performance of the two groups was examined. Three syllable conditions (two- vs. three- vs. four-syllable) were subjected to three within-subjects dependent variables and language background was a between-subject factor to ANOVA. Effect of syllable condition was found significant ($F(1, 2) = 14.61$, $p= 0.000$), but this factor did no interact with language background ($F(1, 2) = 0.86$, $p=0.43$).

Lastly, in order to make a further investigation on whether syllable length had influence on the two language groups in perceiving a specific tonal contrast – minimal pairs (accented on the final position and no pitch accented at presence), three syllable conditions (two-, three- and four- syllable) in minimal pair were subjected to three within-subject dependent factors and language background was between-subject factor. The effect of syllable length in minimal pair was found significant ($F (2, 110) = 26.23$, $p=0.000$) and this factor significantly interacted with language group ($F (2, 110) = 6.978$, $p=0.001$). Post hoc tests comparison were conducted and showed Dutch subjects' perception of minimal pairs in two- syllable was significantly better than in three-syllable and four-syllable conditions (Bonferroni- corrected $p = 0.001$ and $p <0.001$, respectively) but they performed equally poorly in the three- and four-syllable conditions ($p=1.000$). Japanese subjects had significant better discrimination in the two-syllable condition than in the three-syllable condition ($p=0.001$). No significant difference was found between the two-syllable and four syllable conditions ($p=0.237$), and between three-syllable and four-syllable conditions ($p=0.07$).

Two syllable condition

Language	Mean accuracy rate	Std. Deviation	Mean RT (msec)	Std. Deviation
Dutch	91.57%	0.08	767.06	324.04
Japanese	92.94%	0.09	470.06	66.96

Table 3. Results on perception of two-syllable nonwords.

Three syllable condition

Language	Mean accuracy rate	Std. Deviation	Mean RT (msec)	Std. Deviation
Dutch	94.61%	0.07	756.70	342.13
Japanese	97.01%	0.04	439.54	73.77

Table 4. Results on perception of three-syllable nonwords.

Four syllable condition

Language	Mean accuracy rate	Std. Deviation	Mean RT (msec)	Std. Deviation
Dutch	96.99%	0.03	614.90	231.08
Japanese	97.02%	0.04	539.91	62.73

Table 5. Results on perception of four-syllable nonwords.

	LH vs. LH'		LH vs. HL		LH' vs. HL	
Language	Mean accuracy	Std. Deviation	Mean accuracy	Std. Deviation	Mean accuracy	Std. Deviation

	rate		rate		rate	
Dutch	85.86%	0.17	91.95%	0.13	90.93%	0.14
Japanese	87.63%	0.14	93.47%	0.11	93.70%	0.15

Table 6. Results on perception each contrast in two-syllable nonwords.

	LHH vs. LHH’		LHH vs. HLL		LHH vs. LHL	
Language	Mean accuracy rate	Std. Deviation	Mean accuracy rate	Std. Deviation	Mean accuracy rate	Std. Deviation
Dutch	61.93%	0.17	92.79%	0.13	92.56%	0.15
Japanese	70.00%	0.22	92.60%	0.12	97.60%	0.05
	LHH’ vs. HLL		LHH’ vs. LHL		HLL vs. LHL	
Language	Mean accuracy rate	Std. Deviation	Mean accuracy rate	Std. Deviation	Mean accuracy rate	Std. Deviation
Dutch	94.56%	0.15	92.86%	0.13	91.98%	0.11
Japanese	97.53%	0.10	96.73%	0.12	94.20%	0.10

Table 7. Results on perception of each contrast in three-syllable nonwords.

	LHHH vs. LHHH’		LHHH vs. HLLL		LHHH vs. LHLL	
Language	Mean accuracy rate	Std. Deviation	Mean accuracy rate	Std. Deviation	Mean accuracy rate	Std. Deviation
Dutch	60.60%	0.14	96.30%	0.08	95.87%	0.07
Japanese	81.93%	0.09	97.60%	0.05	97.44%	0.09
	LHHH vs. LHHL		LHHH’ vs. HLLL		LHHH’ vs. LHLL	
Language	Mean accuracy rate	Std. Deviation	Mean accuracy rate	Std. Deviation	Mean accuracy rate	Std. Deviation
Dutch	95.09%	0.11	95.13%	0.08	93.53%	0.10
Japanese	96.73%	0.07	97.16%	0.10	96.56%	0.10
	LHHH’ vs. LHHL		HLLL vs. LHLL		HLLL vs. LHHL	
Language	Mean accuracy rate	Std. Deviation	Mean accuracy rate	Std. Deviation	Mean accuracy rate	Std. Deviation

Dutch	94.30%	0.13	92.60%	0.14	95.77%	0.07
Japanese	92.60%	0.13	94.60%	0.13	95.07%	0.09
		LHLL vs. LHHL				
Language	Mean accuracy rate		Std. Deviation			
Dutch	94.05%		0.10			
Japanese	96.73%		0.07			

Table 8. Results on perception of each contrast in three-syllable nonwords.

4.3 Discussion

Dutch and Japanese subjects were required to perform an ABX task based on contrastive tones in different positions in two-, three- and four-syllable conditions. Their performance in perceiving tonal contrasts signaled by pitch position in general was not affected by syllable length. In each syllable condition, Dutch subjects on average made as few errors as Japanese subjects, and both the two groups overall achieved an average accuracy rate above 90%. However, even though Dutch subjects on average performed as well as Japanese subjects, they spent significantly longer time than Japanese subjects to correctly perceive positional contrasts in each syllable condition.

As for subjects' performance in minimal pairs, both groups had poorer discrimination (Dutch group with around 85%, 61%, 60% and Japanese group with around 87%, 70%, 82% in two-, three-, four-syllable conditions, respectively), compared with their performance in other contrasts in which both groups achieved average accuracy of over 90%. Interestingly, Dutch and Japanese subjects only showed significant mean difference in perceiving four-syllable minimal pairs LHHH vs. LHHH'. Japanese subjects achieved significantly better perception than Dutch subjects. In two- and three- syllable condition, no significant mean difference was shown between the two groups in perceiving minimal pairs LH' vs LH and LHH' vs. LHH. Dutch subjects and Japanese subjects had equally poor performance in perceiving the minimal pairs in these two syllable conditions. Moreover, syllable length had an effect on Dutch subjects but not on Japanese subjects in perceiving minimal pairs. Dutch subjects achieved better discrimination in shorter syllable condition (two-syllable) than longer syllable conditions (three- and four-syllable), while Japanese subjects showed better performance between two-syllable and three-syllable condition, but not two-syllable and four-syllable condition. Indeed, Japanese subjects achieved equally good perception in the two- and four-syllable conditions, while Dutch subjects performed equally poorly in longer syllable conditions (three- and four-syllable).

To conclude, the ABX task demonstrated that a) overall, both Dutch and Japanese subjects achieved good perception of all positional contrasts, except for minimal pairs, in each syllable condition (two-, three- and four- syllable). Both groups had poor discrimination in perceiving minimal pairs but Japanese subjects showed significantly better perception than Dutch subjects when listening to minimal pairs in four-syllable condition. b) Even though Dutch subjects achieved accuracy rates as good as Japanese subjects, they spent significantly more time in responding, which suggests they might resort to acoustic residual in echoic memory to make the right decision in discriminating the positional contrasts in different tonal patterns. c) Syllable length in general does not affect perception of positional contrasts in Dutch and Japanese subjects. However, it only had an influence on the perception of a particular tonal contrast- minimal pair between the two groups. Dutch subjects showed poorer perception when the words became longer (three- and four-syllable), whereas Japanese subjects was not affected by the syllable length.

Note that except for the minimal pairs, perception of all the other positional contrasts in each syllable condition was achieved by an average of 90% accuracy rate in Dutch and Japanese subjects. This ceiling effect (over 90% accuracy rate on average) might be due to three possibilities. One is Dutch subjects may map accentual patterns featuring LH' versus H'L to Weak-Strong versus Strong-Weak native stress patterns. Standard deviation of accuracy rate in each tonal contrast in both language groups was statistically small (less than 0.5). This suggests in both groups, subjects' performance was close to each other. However, Dutch subjects spent significantly longer time to make correct reaction and standard deviation of RTs in Dutch subjects was much larger than in Japanese subjects in all the syllable conditions. This suggests that the time Dutch subjects spent in making correct perception varied among individuals. Some might resort to acoustic residual in echoic memory to make correct discrimination. The second possibility is that some Dutch subjects who spent a short time to make correct response might succeed in representing pitch accent marking, or they might use positional marking in stress in native language to perceive the tonal contrasts in different positions. The third possibility is that the ABX task is not demanding enough or "not optimal" (Dupoux et al. 2001:1607). ABX task might allow subjects to rely on alternative acoustic strategies such as paying close attention to acoustic level of representations. Concerning this possibility, a more demanding task was carried out in the next session.

5. Experiment 2 (Sequence recall task)

The second experiment in the study applied a sequence recall task, adapted from the paradigm in Dupoux et al. (2001). Previous research by Dupoux et al. (1997) has revealed that French subjects, different from Spanish subjects, have difficulties in discriminating nonwords that only differed in stress location via speeded ABX tasks. Although a speeded ABX task allowed them to find significant group differences, it did not allow for "a sorting of individual differences" in stress deafness and French subjects' performed better than chance level (50%) (Dupoux et al. 2001: 1606). This might mean either that French subjects resorted to relying on a certain acoustic correlate of stress (e.g., vowel duration[14], higher pitch on the stressed syllable), or that the ABX task was not demanding enough and allowed them to rely on acoustic strategies such as focusing on onset or offset of the syllable. Given this concern, Dupoux et al. (2001) developed a new paradigm based on memory load - the sequence recall paradigm. Generally speaking, the paradigm involves a memory task. Subjects listen to two words combined in orders (such as words A-B or B-A) in a various word sequence (such as three-word sequence: A-B-B) and are required to recall the word order they have listened to in the sequence. The longer the word sequence, the higher the memory load is imposed on subjects. This more demanding paradigm tried to eliminate subjects' employment of acoustic strategies and aimed to examine speech perception at a "more abstract level", a phonological processing level (Dupoux et al. 2001:1615). Details of this paradigm were presented in the following sections.

5.1 Method

5.1.1 Subjects

Subjects were the same as in Experiment 1.

5.1.2 Stimuli

The sequence-recall paradigm so far has been applied in investigating bisyllabic words, with testified robustness (Dupoux et al. 2001, 2008, 2010; Utsugi et al. 2010). Accordingly, the current study used bisyllabic words. Bisyllabic nonword "nono" was constructed in three possible accent conditions: H'L (initial syllable accented), L-H' (final syllable accented) and L-H (unaccented), which differ in the presence of pitch accent (HL pitch fall) and in the position where the pitch accent falls. The three tonal patterns result in three contrasts: H-L vs. LH', H'L vs. LH and LH' vs. LH as illustrated in Table 9. As stated before, LH' vs LH is

[14] Vowel duration in the stressed syllable was 20 ms longer than unstressed syllable. Maximum f0 was 45.3 Hz in the stressed syllable higher than unstressed syllable (Dupoux et al. 2001).

minimal pair that only differ in the final tone. The H tone in LH' (accented on the final position) is slightly higher than the H tone in LH (unaccented), but the H tone in LH is still much higher than the L tone in the tonal pattern.

Positional Contrast	Surface tonal patterns	
	Token A	Token B
Initial accented vs. Final accented	H'L	LH'
Initial accented vs. Unaccented	H'L	LH
Final accented vs. Unaccented	LH'	LH

Table 9. Contrasts of tonal patterns used in the experiment**.**

Tokens "nono" were clearly articulated in the same carrier sentence as in Experiment 1 and recorded by a trained Japanese female phonetician (not the same one as in Experiment 1 to eliminate confounding) and were checked by another phonetician. Both the phoneticians are native speakers of Tokyo Japanese. Each token was digitized at 16 KHz at 16 bits. Maximal and minimal f0 values were controlled and peak intensities of the H accented tones were set to the same value. Any tokens containing lengthened vowels and/or glottalized final syllables were eliminated. The average duration of the tokens was 96.1 ms. The stimuli are attached in the Appendix.

5.1.3 Procedure

The experiment was modified on the paradigm used by Dupoux et al. (2001). The experiment consisted of three parts, each part containing one positional contrast. The order of the parts was counterbalanced. Each part consisted of three sessions: a learning session, a practice session and a test session. Subjects could take a five-minute break after each part.

In learning session, all the subjects were first told that they would learn two new words in a foreign language. They could listen to various tokens of the two words by pressing button A and button B as many times as they wanted so that they could make sure of associating the buttons with the corresponding tokens. For a given contrast, for instance, nonword "nono" in the H'L tonal pattern was associated with button A and "nono" in the LH' tonal pattern was associated with button B. Pressing a button resulted in playing a token.

After the subjects were comfortable with the button - token association, they proceeded to the practice session. The practice session consisted of two blocks. Block one contained 18 trials. Subjects would listen to a token and then were required to identify the token by pressing the associated button A or B. They were allowed to press the button after the word "Press" showed on the screen. They would receive a feedback "correct" or "incorrect" displaying for 800 ms. on the computer screen after their response. A 1500 ms. pause separated each response from the next trial. In Dupoux et al. (2001), a five consecutive correct response was set as a criterion, that is, subjects were allowed to proceed to block two after they achieved five correct responses in a row. However, during our pilot tests, it was observed that in the contrast of L-H' vs L-H, some subjects (6 out of 10 Dutch subjects) failed to achieve this criterion after trying several times in half an hour, which prevented them from proceeding to the next block. Given this concern, the criterion in block one was modified to a success of 12 correct responses (not in a row) among 18 trials in total. Subjects would not proceed to block two unless they achieved this criterion. Block two consisted of 8 trials. In block two, subjects would listen to the two tokens in various combination orders in two-word (e.g., A-B), three-word (e.g., A-B-A) and four word (e.g., A-B-A-B) sequences. They were then required to press the associated buttons according to the order of tokens they heard. For instance, if a subject heard token A- token A, he had to press the button A, A in order. Subjects were only allowed to press the buttons after a "Press" showed up on the screen, in order to prevent them pressing the buttons during listening. A feedback of "correct" or "incorrect" was presented and displayed for 800 ms. on the screen after their response. They were told that they could do practice in block two as many times as they wanted, until they made sure they were able to proceed to the test session. Each trial in block two lasted for 1500 ms and each token in one trial was separated by an interval of 80 ms. 8 trials contained two trials for each sequence and all trials were counterbalanced.

The test session consisted of three blocks, each block containing 8 trials. Tokens in two-word, three-word and four-word sequence were presented in blocks one, two and three, respectively. Token A and B in two-word sequence resulted in four combination orders: A-B, A-A, B-B and B-A. Each of the four orders were presented twice in block one. The two tokens in three-word sequence resulted in eight order combinations: A-A-A, A-A-B, A-B-A, A-B-B, B-B-B, B-B-A, B-A-B and B-A-A. Each combination was presented once in block two. The two tokens in four-word sequence resulted in 16 order combinations, but eight order combinations with an A to B or B to A variation were selected: A-A-B-A, A-A-B-B, A-B-A-A, A-B-B-A, B-B-A-B, B-B-A-A, B-A-B-B and B-A-A-B. Each order combination was presented once in block three.

The overall stimuli design in test session was constructed as: contrast (H'L-LH', H'L – LH, LH' - LH) x sequence length (two- vs. three- vs. four-word) x word order (eight combination), resulting in 72 (3 x 3 x 8) trials in total. Each trial lasted for 1500 ms. One trial would proceed to the next if the subject failed to react within this time pressure. A "beep" sound and a mark "OK" appeared after each trial to avoid subjects using echoic memory. No feedback was shown on the screen after subjects' response. Note that five-word and six-word sequence length were not applied in this study due to the finding in Dupoux et al. (2001) that subjects had chance performance in these two sequence lengths.

The experiment lasted 15 minutes in total and was conducted on a laptop with a headset and a two-button button box, programmed in ZEP (Veenker 2012). Responses were recorded using ZEP software. Dutch subjects participated in the experiment in a sound proof phonetic lab in Utrecht Institute of Linguistics OTS in Utrecht University. Japanese subjects were tested in a sound proof room or a quiet room in University of Tokyo, Tokyo University of Foreign Studies and The University of Electro-Communications.

5.2 Results

Recorded responses were transformed into accuracy rates. Accuracy rates in each contrast and in each block (two-, three- and four-word sequence) in each subject were analyzed and transformed to z-score. Based on van Selst and Jolicoeur (1994), the subject whose accuracy rate was beyond the range of (-2.246, 2.246) was regarded as an outlier. In the data of H’L vs. LH’, two subjects (a z-score: -3.26, -2.73 respectively) in Dutch group and one subject (a z-score: -2.61) in Japanese group were eliminated. In the data of H’L vs. LH, data of four subjects (a z-score: -3.09, -2.62, -2.73, -3.26) in Dutch group and one subject (a z-score: -2.51) in Japanese group were eliminated. In the data of LH’ vs. LH contrast, two subjects (both z-scores: -2.74) in Dutch group and one subject (a z-score: -2.74) in Japanese group were eliminated.

To start with, whether Dutch and Japanese subjects were influenced by positional contrast and/or by sequence length was examined. Contrast (H’L vs. LH’, H’L vs. LH, LH’ vs. LH) and Sequence Length (two vs. three vs. four) were subjected to two within-subject dependent variables in ANOVA. Language background was one between-subject factor (Dutch vs. Japanese). No effect of contrast or length was observed (contrast: $F(2, 98) = 0.73$, $p=0.49$); length: $F(2, 98) = 1.13$, $p=0.28$). The factor length did not interact with language background ($F(2, 98) = 1.00$, $p=0.37$), whereas the factor contrast was found interaction with language background ($F(2, 98) = 3.24$, $p=0.04$).

Since the interaction between contrast and language background was observed, secondly, a further investigation was conducted into which contrast affected Dutch and Japanese subjects performance. Contrast (H’L vs. LH’, H’L vs. LH, LH’ vs. LH) was subjected to one within-subject dependent variable and language background was between-subject factor in ANOVA. Accuracy rates and standard deviation of the two groups in each contrast were displayed in Table 10 and Figure 2. Dutch subjects made significantly more errors than Japanese subjects in perceiving the minimal pair LH’ vs. LH contrast ($F(1, 1) = 6.20$, $p=0.016$). Individual performance variation in Dutch subjects was also observed larger than that in Japanese subjects when perceiving the LH’ vs. LH contrast. No significant mean difference was observed in H’L vs. LH’ and H’L vs. LH between Dutch and Japanese subjects, respectively (H'L vs. LH': F (1, 1) = 1.75, p=0.191; H'L vs. LH: F (1, 1) = 0.67, p=0.417).

Given the findings discussed above that only the LH' vs. LH contrast resulted in significant mean difference between the two groups, a separate test was carried out to examine whether sequence length affected subjects' performance within this contrast particularly. Accuracy rates and standard deviation are reported in Table 11. Sequence length (two-, three- and four-word sequence) was subjected to one within-subject dependent factor and language background was between-subject factor by an ANOVA. In perceiving LH' vs. LH, Dutch subjects made significantly more errors than Japanese subjects in all the sequence length conditions (two-word sequence: F (1, 2) = 3.94, p=0.047; three-word sequence: F (1, 2) = 4.45, p=0.040; four-word sequence: F (1, 2) = 6.68, p=0.012). However, no effect of sequence length was found (F (2, 104) = 2.54, p=0.083) on each group, and this factor did not interact with language background (F (2, 104) = 1.10, p=0.336). This indicated that, regardless of sequence length, it was the contrast (LH' vs. LH) that affected the distinguished performance between Dutch and Japanese subjects.

	H'L vs LH'		H'L vs LH		LH' vs LH	
Language	Mean accuracy rate	Std. Deviation	Mean accuracy rate	Std. Deviation	Mean accuracy rate	Std. Deviation
Dutch	77.79%	0.21	83.91%	0.21	77.66%	0.24
Japanese	86.17%	0.16	88.83%	0.12	94.48%	0.07

Table 10. Dutch and Japanese subjects' performance in each contrast.

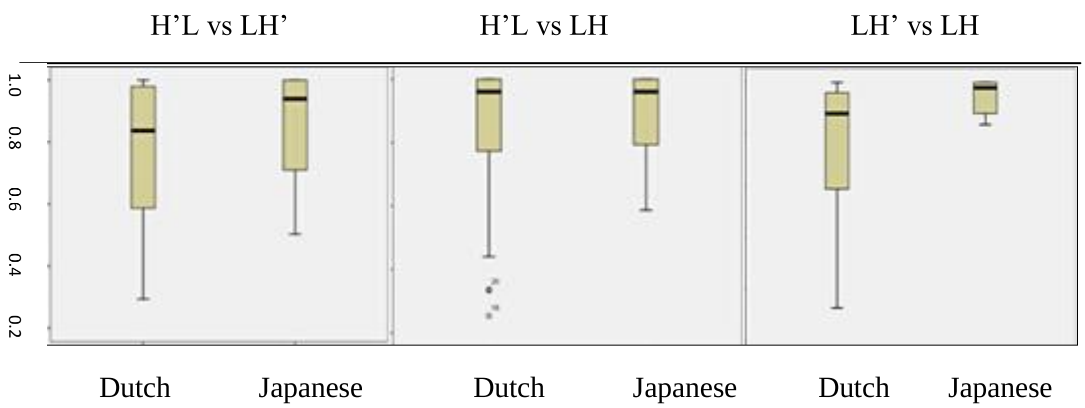

Figure 2. Dutch and Japanese subjects' performance in each contrast.

LH' vs. LH						
	2-word sequence		3-word sequence		4-word sequence	
Language	Mean accuracy rate	Std. Deviation	Mean accuracy rate	Std. Deviation	Mean accuracy rate	Std. Deviation
Dutch	80.38%	0.23	84.58%	0.19	77.40%	0.24
Japanese	93.86%	0.09	95.71%	0.06	93.86%	0.09

Table 11. Dutch and Japanese subjects' performance in LH' vs LH regarding sequence length.

5.3 Discussion

Dutch and Japanese subjects were required to recall bisyllabic tokens of positional contrasts in various order combinations and sequence lengths in Experiment 2. This sequence-recall task demonstrated that contrast, but not sequence length, in general affected the performance of both groups in perceiving tonal contrasts. It was found that Dutch subjects had more difficulties than Japanese subjects in perceiving the minimal pair LH' vs. LH, whereas both groups did not have problems in perceiving the other contrasts H'L vs. LH, H'L vs. LH'.

Furthermore, in perception of the minimal pair, sequence length had no influence on performance within or between the two groups. It was only the contrast that resulted in significant perceptual difficulties in Dutch subjects, regardless of sequence length. Dutch subjects always made significantly more errors than Japanese subjects in each sequence length condition. The longer the word sequence, the higher memory load was imposed on subjects. However, Dutch subjects' perception did not change with the imposed memory load. They had equally poor performance even with a low memory load (two-word sequence length), whereas Japanese subjects remained good perception with increasing memory load. In the findings of Dupoux et al. (2001), French listeners had no detectable problem detecting stress location in short sequence length (two- and three-word sequence) but had trouble with long sequence lengths (four-, five-, six-sequence[15]); they showed stress deafness in short sequence length when phonetic variability (tokens they were recorded by two speakers) was incorporated with memory load. The stress deafness observed in French listeners occurred only in the condition imposed with memory load and phonetic variability (Dupoux et al. 2001), whereas Dutch subjects' deafness in the tonal contrast in final position was found regardless of phonetic variability and memory load.

[15] In the findings of Dupoux et al. (2001), French subjects had chance performance in five- and six-word sequence lengths.

The differences between the performance of Dutch and Japanese subjects in perceiving the specific final position might be due to an effect of non-native vs. native perception. Yet, the perceptual failure observed in Dutch subjects suggested that they encountered problems in processing tonal differences when the two tones differed slightly in phonetic and phonological features in the word final position. Phonetically speaking, the H tone in the final accented position in a word is slightly higher than the H tone in the final position when the word is unaccented. At the phonological level, the slight surface tonal differences in the final position are realized due to underlying accentual patterns (Kawahara 2015). Accentual patterns differ in whether the pitch accent marking occurs; and if the accent occurs, accentual patterns differ in the location of the pitch accent. In this sense, the failure of Dutch subjects in detecting contrasts of tonal positions in minimal pairs implies that Dutch subjects had difficulty in processing pitch accent marking in final position at phonological level. The employment of positional marking in stress seemed not to equip Dutch listeners to process the specific tonal position at phonological level. Indeed, Nooteboom & Doodeman (1985), via gating experiment (Grosjean 1980), examined the Dutch listeners' perception of stressed syllable in the different positions in accented condition. They found that Dutch listeners showed an advantage in perceiving stress marking when the stress was on word medial position while the advantage of stress marking was weakly perceived in word final position. Another study conducted by Jongenburger & van Heuven (1995a; see also Jongenburger 1996) examined stress perception in word recognition in unaccented condition and found that Dutch listeners were able to detect and judge stress only on the initial syllable but not on the final syllable in a word when recognizing real words and nonwords. Another account for the different performance between Dutch and Japanese subjects in processing tonal contrasts in final position might be the usage of lexical pitch. Tonal difference in final position is signaled by the presence of pitch accent marking. Dutch uses pitch as an acoustic correlate of stress, while pitch is exclusively used in Japanese in marking positions to differentiate lexical contrasts. Cutler & Otake (1999; also see Otake et al. 1993) found that the marking of pitch accent in morae was fully employed by Japanese listeners for word recognition.

In sum, Dutch subjects were able to perceive positional contrasts in bisyllabic nonwords, as well as Japanese subjects, except for the minimal pair (LH' vs LH). Their performance remained equally poor in perceiving the minimal pair regardless of memory load, while Japanese subjects achieved good perception in all memory load conditions. The perceptual deafness in Dutch subjects in the specific final position occurred even without phonetic variability, different from French subjects' deafness for stress that was only found in a more demanding condition with memory load and phonetic variability (Dupoux et al. 2001). This indicated that Dutch subjects encountered difficulties in processing the presence of pitch accent marking in the final position in particular, though they were able to detect tonal contrasts in other positions.

6. General discussion

Experiment 1 (ABX task) and Experiment 2 (sequence-recall task) showed similarities but also differences in the findings. Both experiments examined positional contrasts but Experiment 1 focused on examining positional contrasts in different syllable conditions (two-, three- and four-syllable nonwords) at the acoustic processing level while Experiment 2 centered on pitch position processing at the phonological level via investigating positional contrasts in bisyllabic nonwords. In general, the two experiments demonstrated that Dutch subjects processed contrasts of pitch position acoustically, and Japanese subjects processed contrasts of pitch positions at the acoustic level and phonological level.

Both experiments revealed that Dutch subjects were able to detect tonal contrasts in different positions as well as Japanese subjects, except for a specific final position. Dutch subjects succeeded in detecting syntagmatic tonal differences (position in a word), that is, the pitch-accented syllable distinguished itself from the unaccented syllables surrounding it, for example, the accented syllable in the initial position in H'LLL vs. in the second position in LH'LL. Additionally, Dutch subjects were able to perceive syntagmatic tonal variation in sequence in opposed positions, for example, H'-L in final position in LHH'L vs. H'-L in medial position in LH'LL. Their success in perceiving syntagmatic tone oppositions at the acoustic and phonological levels may due to two factors. One is that positional marking of stress in Dutch may facilitate detection of syntagmatic tonal contrasts. Previous studies have shown Dutch listeners exploit stress position for word recognition (Cutler & Otake 1990; Cutler 2001; van Heuven 1986; Nooteboom & Doodeman 1985; Jongenburger & van Heuven 1995a). Cutler (2001) investigated Dutch listeners' perception of stress location in twelve minimal pairs such as *CAnon* ("norm") vs. *kaNON* ("gun") when recognizing words. She found that the minimal pairs which differed in suprasegmental information - stress position were not processed as homophones in activating lexicon. The employment of stress position "constrained the activation of lexical representations" (Cutler 2001:186), that is, when the Dutch subjects listened to the early suprasegmental information - stressed vs. unstressed initial syllable, the corresponding lexical representation of the word was activated. For instance, the stressed syllable *VOOR* (in *VOORnaam*"first name") only activated the lexical representation of *VOORnaam* while the unstressed syllable *voor* (in *voorNAAM* "respectable") only activated the lexical representation of *voorNAAM*. The other possible factor in the success in Dutch subjects in detecting pitch positions may be due to mapping tonal contrasts onto intonation patterns. According to the autosegmental-metrical (AM) framework (Pierrehumbert 1980; Ladd 1996), Dutch has eight types of nuclear pitch accent, H*[16], L*,

[16] Asterisk (*) here stands for nuclear accent.

H*L, L*H, L*![17]HL, !H*L, L*HL (delayed fall) and H*!H (vocative chant) (Gussenhoven 2005). Dutch subjects may form the tonal patterns into nuclear pitch accent contours in Dutch intonation system, for instance they may form tonal pattern HLL as a nuclear pitch accent contour and then map it into H*L or !H*L, map LHH' into L*H, LHHL into L*HL.

In addition to the similar findings in the two experiments, different results were presented as well in perception of tonal contrasts in the specific final position in minimal pairs (final syllable accented vs. unaccented). The minimal pair of accented vs. unaccented final syllable have identical surface tonal patterns but differ slightly on the final tone. Phonologically speaking, the tonal difference on the final position is due to the presence vs. absence of pitch accent marking. Acoustically speaking, the H tone in the final syllable in an unaccented word was slightly lower (but still higher than any tones in the other syllables) than the H tone in the final accented syllable. Experiment 1 examined perception of pitch contrasts in this final position in two-, three- and four-syllable conditions. It revealed that Dutch subjects only showed significantly poorer performance than Japanese subjects in the four-syllable condition. This indicated that Dutch subjects were able to detect tonal differences contrastive on the final position when the syllable was short (LH' vs. LH, LHH' vs. LHH). One account for their success in short syllable conditions is that they may rely on acoustic strategies such as paying close attention to the acoustic cue in the final position. However, when the final tone differences occurred in a longer syllable condition, employment of acoustic residual cues may not be reliable enough. Indeed, though Experiment 1 did not show Dutch subjects had perceptual problems in two-syllable words, Experiment 2, incorporating a memory task, demonstrated that Dutch subjects had difficulties in detecting tonal differences in two-syllable words at phonological level whereas Japanese subjects had good perception. The task in Experiment 2 was more demanding than in Experiment 1 in that subjects had to respond to the combined order of this particular contrast in two-, three- and four-word sequence. Even in the short sequence length (two-word sequence), Dutch subjects demonstrated perceptual difficulties and their poor discrimination remained in all sequence length conditions.

The misperception on the specific final position in Dutch subjects can be accounted for by several reasons. One is that Dutch listeners show perceptual bias on the final position in stress marking (van Heuven 1986; Nooteboom & Doodeman 1985; Jongenburger & van Heuven 1995a), which may influence their perception of pitch positional contrasts. Nooteboom & Doodeman (1985) found that Dutch listeners showed the most advantage in recognizing words when the stressed syllable was in word medial position in accented marking condition, while the effect of the stressed syllable in word final position was weak.

[17] Exclamation mark (!) represents a downstep accent. For instance, the L tone with "!" is lower than the previous L tone.

To tease apart the interference with accent marking, van Heuven (1984, 1986) investigated the effect of stress position without accent marking for word recognition. He found in a gating experiment (Grosjean 1980) that Dutch listeners assumed the first syllable was stressed, irrespective of the actual stress position in the word. Likewise, Jongenburger & van Heuven (1995a) found that when recognizing words, Dutch listeners detected and judged stress only on the initial syllable but not on the final syllable in a word. This perceptual bias may be due to the asymmetrical behavior of affixes in Dutch (van Heuven 1986; Jongenburger & van Heuven 1995a). Stress in Dutch stem morphemes is often shifted under the influence of a suffix, which may either bear the stress itself, or attract the stress to a syllable one or two position before the suffix. Prefixes and affixes in general, on the contrary, never affect the stress to "shift" towards the beginning of a word, and are "typically unstressable themselves" (van Heuven 1986: 15).

Another concern for the misperception on the final position in Dutch subjects may be due to the employment of linguistic pitch. The tonal contrast on this specific final position is caused exclusively by the pitch accent marking *phonologically* (Kawahara 2015). Since the duration of each syllable was controlled to the same and no long vowels were used, pitch was the exclusively acoustic cue to rely on. In Dutch, a stressed syllable is realized by longer vowel duration, vowel quality, louder intensity or higher pitch (Quené 1992; Sluijter & van Heuven 1996a, 1996). However, the effect of pitch for cuing lexical stress is not as salient as other cues, like duration. Quené (1992) found that vowel duration was a salient cue for Dutch speakers to segment words, while pitch differences were used as acoustic correlate of stress. Sluijter & van Heuven (1996a) investigated the effect of f0, intensity and duration for cuing stress in stress minimal pairs in Dutch. They found f0 movement differentiated only between accented and unaccented syllables but not between stressed and unstressed. Likewise, in a production study, Sluijter & van Heuven (1996) examined the effect of Dutch lexical stress on overall intensity, spectral balance, and vowel quality and duration without interference from accent condition. They found that vowel quality and overall intensity were weak cues for lexical stress; duration was the most salient cue while spectral balance was also found close in strength to vowel duration. These findings indicated that pitch in Dutch is one of the acoustic correlates of lexical stress, but not a salient one, compared with other acoustic cues such as vowel quality and duration. Dutch listeners may be able to exploit pitch differences on the final position in short words at the acoustic level. However, the extent to which they exploit pitch differences to track positional tones did not equip them to process tonal contrasts in the specific final position at the phonological level.

The current study examined predictions for suprasegmentals made by different types of speech perception models. First, models relying on phonemic status and phonetic features (Best 1995; Best et al. 1988, 2001; Best & Tyler 2007) predicted that tonal contrasts involving HL vs. LH combination would be assimilated to two categories and yield good discrimination (see Section 2.3.3.1). Indeed, pairs of tonal contrasts such as LH'L-LHH' may have been a Two Category (TC) assimilation pair for Dutch listeners. According to PAM, TC occurs when two non-native phones are assimilated to two separate native categories. In this sense, LH'L may be assimilated to nuclear pitch accent category L*HL or L*!HL while LHH' may be assimilated to L*H. Contrastive tonal pair LH' – H'L (or LH – H'L) may be a TC assimilation pair as well in that LH' – H'L may be assimilated to Weak-Strong and Strong-Weak, respectively in stress patterns (one syllable marked stronger and more prominent than the other), or may be assimilated to nuclear tone categories L*H and H*L respectively (Gussenhoven 2005). For contrastive pair such as LHH'L – LH'LL, if according to PAM and given nuclear tone categories in Dutch, they may have been assimilated to L*HL or L*!HL intonation category, regarded as one Single Category (SC). When two non-native phones form a SC assimilation pair, poor discrimination is expected, which contradicts the finding that Dutch listeners achieved good discrimination in LHH'L – LH'LL. When it comes to the perception of tonal contrasts specifically in the final position (H' vs. H), PAM predicted that the pair (LH' - LH) may be assimilated into the same L*H nuclear tone category and then yield poor discrimination, which supported the finding that they had problems in perceiving the minimal pair at the phonological level. However, an alternative prediction made in Section 2.3.3.1 predicted good discrimination for minimal pairs. According to PAM, if the non-native phone did not exist in the native perceptual system, it would not be assimilated to any existing category but establish a new category resulting in good discrimination. Dutch has no lexically contrastive pitch and the positional pitch contrast is dissimilar to any existing native category in Dutch word prosody, which should have been easy for Dutch subjects to discriminate contrastive pitch in the final position, which is opposed to the finding in the current study. Alternatively, a word without pitch accent marking did not resemble stress marking in Dutch word prosody since a syllable in a word is always stressed, and thus words without marking seemed not to exist in Dutch. This would result in good discrimination, which contradicts the finding.

Second, unlike models based on assimilation/dissimilation of phonemic details, models proposed in terms of the effect of native phonological representation (Brown 1998, 2000) predicted good discrimination between pitch positional contrasts but impairment in perceiving minimal pairs (see Section 2.3.3.2). Indeed, nuclear tone types in Dutch intonation prosody feature L*H, H*L, L*HL and L*!HL (Gussenhoven 2005), which may equip Dutch listeners to form tonal patterns such as LH'L, LHH'L, LHLL or LH' (LH) and H'L tonal combination as a nuclear pitch contour. The presence of these tonal contrasts at the intonation level in Dutch may help Dutch listeners perceive non-native positional contrasts. Another concern is the positional tones featuring H'L and LH' (LH) may also be associated with stress patterns (i.e., Strong-Weak, Weak-Strong, respectively) in Dutch. The presence of tonal features in native language, either at the intonational level or in word prosody, may allow Dutch listeners to perceive positional tones. On the other hand, Brown's model (1998, 2000) predicted that Dutch listeners would encounter difficulties in perceiving minimal pair contrasts, which was supported by the current findings. The minimal pair examined in the study differed in the tone in the final position: the H tone in the accented final syllable was slightly higher than that without pitch accented. The tonal patterns in Tokyo Japanese are mapped from underlying accentual patterns where pitch accent occurs. Thus tonal positions in the minimal pair indicate the contrastive condition of the absence vs. presence of pitch accent (HL pitch fall). However, there is no equivalence of unaccentedness in Dutch but there is always stress in Dutch words. Additionally, Dutch listeners may look for a pitch peak instead of a HL pitch fall in that a stressed syllable is realized with higher pitch peak than unstressed syllable (Sluijter & van Heuven 1996). Dutch listeners may map both accented syllable and unaccented syllable to Dutch word final stress. This makes it easy to distinguish unaccented position from most positions, but not from the final accented position. Given these concerns, the lack of unaccentedness in Dutch may result in perceptual problems for this specific tonal position.

Lastly, according to the Feature Hypothesis, Dutch subjects should encounter more difficulties in perceiving positional tone contrast than Japanese subjects, which was opposed to the findings. This model proposed that the more the linguistic pitch was used, the better perceptual performance would be expected. However, the results showed Dutch subjects performed as well as Japanese listeners in most tonal contrasts in different positions. Dutch subjects demonstrated a deficit only in perceiving minimal pairs. Indeed, the hypothesis supported the finding in perception of minimal pairs in that the specific final position was signaled by a HL pitch fall, which is not used to signal lexical contrasts in Dutch word prosody. Lexical pitch in Dutch is one of the acoustic correlates of stress, but is less salient than vowel quality and duration (Quené 1992; Sluijter & van Heuven 1996) while it is exclusively used in Japanese to contrast word meanings.

To conclude, the present findings demonstrated that Dutch listeners were partly "deaf" to non-native pitch positions. The usage of positional marking in stress may equip Dutch listeners to perceive most non-native contrasts of tonal positions (e.g., H tone in syntagmatic positions or tonal sequence H'L in syntagmatic positions). However, they failed to detect a specific positional contrast (in minimal pairs) due to the absence of pitch accent marking versus pitch accent on the final position, whereas Japanese listeners showed no problems in detecting all the contrasts of tonal positions. The findings carry some implications for theoretical models with respect to non-native suprasegmental perception. Models based on abstract phoneme such as PAM (Best 1995; Best et al. 1988, 2001; Best & Tyler 2007) supported most of the findings in the current study, but failed to account for Dutch listeners' success in discriminating LH'LL-LHH'L contrast. According to PAM, this pair may be assimilated to a single category in Dutch native intonation type such as L*HL or L*!HL and thus result in poor discrimination, which contradicted the finding in the current study. Feature

Hypothesis failed to predict that Dutch listeners had no problem in discriminating most of the contrasts of pitch positions, but supported the findings in perception of the minimal pairs. The model based on phonological representation (Brown 1998, 2000), on the other hand successfully predicted the findings in the current study. The presence of a phonological representation, either in nuclear pitch accent patterns or in word stress patterns, may equip Dutch listeners to perceive most non-native positional contrasts. However, the lack of unaccentedness in Dutch caused a perceptual deficit in tonal contrasts in the final position.

Besides, except for minimal pairs, the ABX task in the current study showed Dutch subjects were able to process tonal position contrasts in bisyllabic words at the acoustic level, while the sequence-recall task revealed that they failed in processing tonal position contrasts in bisyllabic words at the phonological level. The results have limitations in that the memory load incorporated in the second task might not be demanding enough. In the study of stress "deafness" in French listeners, Dupoux et al. (2001, 2008) examined the sequence-recall paradigm in different conditions. They set a criterion of the highest memory load at five- or six-word sequence length and found French subjects performed by chance level in these two conditions, which was not applied in the present study. In addition, though they found the paradigm robust using the material recorded by one female speaker, they incorporated phonetic variability in another condition, that is, the experimental materials were recorded by two female speakers and one male speaker. The present study used the tokens recorded by one female speaker, which might limit the study to examine the perception of H'L vs. LH and H'L vs. LH' at the phonological level. Yet, although no phonetic variability was involved, Dutch listeners still exhibited perceptual "deafness" to the unaccented final position vs. accented final position, irrespective of memory load, which may indicate the influence of the lack of unaccentedness in word prosody in Dutch. This is different from the findings in Dupoux et al. (1997, 2001) that French listeners had no problems detecting stress position in low memory load conditions unless phonetic variability was incorporated. The finding that the lack of unaccentedness in positional marking can cause impairment in perception of tonal contrasts in the final position requires more work to be done in the future. Whether such specific positional "deafness" can be generalized calls for investigating listeners whose native language do not use positional marking at the word prosody such as French or Seoul Korean. Moreover, whether Dutch listeners are "deaf" to positional pitch contrasts is informative to investigate at mapping between the contrasts in novel words and meanings in future study.

References

Abramson, A. S. 1976. Static and dynamic acoustic cues in distinctive tones. Journal of the Acoustical Society of America 59, S42 (A).

Abramson, A. S. 1979. The noncategorical perception of tone categories in Thai, in *Frontiers of Speech Communication Research, eds by B. Lindblom and S. Öhman. Academic, London*: 127–134.

Baayen, R. H., Piepenbrock, R., & Rijn, H. van. 1993. *The CELEX lexical database* (CD-ROM). Linguistic Data Consortium, University of Pennsylvania, Philadelphia, PA.

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

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

Best, C. & Tyler, M. 2007. Nonnative and second-language speech perception: Commonalities and complementarities. In Ocke-Schwen Bohn & Murray Munro (Eds.), *Second-language speech learning: The role of language experience in speech perception and production. A festschrift in honor of James E. Flege* (pp. 13-34). Amsterdam: John Benjamins. Bohn,

Best, C. T., McRoberts, G. W., & Goodell, E. 2001. Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listeners' native phonological system. *Journal of the Acoustical Society of America*, 109, 775–794.

Best, C. T., McRoberts, G. W., & Sithole, M. N. 1988. Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by English-speaking adults and infants. *Journal of Experimental Psychology: Human Perception and Performance*, 14, 345–360.

Blicher, D.L., Diehl, R.L., & Cohen, L. B. 1990. Effects of syllable duration on the perception of the Mandarin Tone 2/Tone 3 distinction: evidence of auditory enhancement. *Journal of Phonetics* 18(1): 37-49.

Bluhme, H & Burr, R. 1971. An audio-visual display of pitch for teaching Chinese tones, *Studies in Linguistics* 22: 51-57.

Brown, C. 1998. The role of the L1 grammar in the L2 acquisition of segmental structure. Second Language Research, 14, 136–193.

Brown, C. 2000. The interrelation between speech perception and phonological acquisition from infant to adult. In J. Archibald (Ed.), Second language acquisition and linguistic theory (pp. 4–63). Oxford: Blackwell.

Brown, C. & Matthews, J. 1997. The role of feature geometry in the development of phonemic contrasts. In S. J. Hannahs and M. Young-Scholten (eds.), *Generative Studies in the Acquisition of Phonology*, pp. 67-112. Amsterdam: John Benjamins Publishing Company.

Brown, C. & Matthews, J. 1993. The acquisition of segmental structure. *McGill Working*

Papers in Linguistics 9: 46-76.

Burnham, D., Francis, E., Webster, D., Luksaneeyanawin, S., Attapaiboon, C., Lacerda, F., & Keller, P. (1996, October). Perception of lexical tone across languages: Evidence for a linguistic mode of processing. In Spoken Language, 1996. ICSLP 96. Proceedings., Fourth International Conference on (Vol. 4, pp. 2514-2517). IEEE.

Burnham, De & Francis, E. 1997. The role of linguistic experience in the perception of Thai tones. In Arthur S. Abramson (Ed.), *Southeast Asian linguistic studies in honour of Vichin Panupong*: 29–46. Bangkok: Chulalongkorn University Press.

Burnham, D., Kirkwood, K., Luksaneeyanawin, S., & Pansottee, S. 1992. Perception of Central Thai tones and segments by Thai and Australian adults. In *Pan-Asiatic Linguistics: Proceedings of the Third International Symposium on Language and Linguistics:* 546-560. Bangkok: Chulalongkorn University Press.

Bybee, J., Charkrabopti, P., Jung, D. & Scheibman, J. 1998. Prosody and segmental effect some paths of evolution for word stress. *Studies in Language* 22: 267-314.

Chan, S., Chuang, C.-K. & Wang, W. S.-Y. 1975. Cross-linguistic study of categorical perception for lexical tone. *Journal of the Acoustical Society of America* 58: S119.

Chen, A. & Kager, R. W. J. 2012. Universals and uniqueness in lexical tone pereption. *Proceedings of Tokyo Conference on Psycholinguistics* 2011: 55-77.

Cutler, A. 1986. Forbear is a homophone: Lexical prosody does not constrain lexical access. Language and Speech, 29, 201–220.

Cutler, A., & van Donselaar, W. 2001. Voornaam is not (really) a homophone: Lexical prosody and lexical access in Dutch. Language and Speech, 44, 171–195.

Cutler, A., & Otake, T. 1999. Pitch accent in spoken word recognition in Japanese. Journal of the Acoustical Society of America, 105, 1977–1988

Dupoux, E., Pallier, C., Sebastián-Gallés, N., & Mehler, J. 1997. A destressing "deafness" in French? Journal of Memory and Language, 36, 406-421.

Dupoux, E., Peperkamp, S., Sebstian-Galles, N. 2001. A robust method to study stress “deafness”. *Journal of the Acoustical Society of America* 110: 1606-1618.

Dupoux, E., Sebastián-Gallés N., Navarrete, E., & Peperkamp, S. 2008. Persistent stress ‘deafness’: The case of French learners of Spanish. Cognition, 106, 682-706.

Eefting, W.Z.F. 1991. The effect of information value and accentuation on the duration of utch wods, syllables, and segments. Journal of the Acoustical Society of America 89, 412-424.

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

Francis, A. L., C, Valter, L. Ma, & F. M. Kimberly. 2008. Perceptual learning of Cantonese lexical tones by tonal and non-tonal language speakers. *Journal of Phonetics*, 36, 268-294.

Gandour, J. 1983. Tone perception in Far Eastern languages. *Journal of Phonetics*, 11, 49-175.

Gandour, Jackson & Harshman, Richard. 1978. Crosslanguage differences in tone perception: a multidimensional scaling investigation. *Language and Speech*, 21, 1-33. 12

Glottfried, T. L & Suiter, T. L. 1997. Effect of linguistic experience on the identification of Mandarin Chinese vowels and tones. *Journal of Phonetics 25*: 207-231.

Guion, S. G. 2004. Training English and Chinese listeners to perceive Thai tones: A preliminary report. Language Learning, 54, 681-712.

Grosjean, F. 1980. Spoken word recognition and the gating paradigm. *Perception and Psychophysics 28, 267-283.*

Goto, H. 1971. Auditory perception by normal Japanese adults of the sounds "l" and "r". *Neuropsychologia, 9,* 317-323.

Gussenhoven, C. 2005. Transcription of Dutch Intonation. In S-A, Jun (ed.), *Prosodic typology and transcription: A unified approach*. Oxford: OU Press, 118-145

Hallé, P. A., Chang, Y, & Best, C. T. 2004. Identification and discrimination of Mandarin Chinese tones by Mandarin Chinese vs. French listeners. *Journal of Phonetics*, 32, 395-421.

Haraguchi, S. 1977. *The Tone Pattern of Japanese: An autosegmental Theory of Tonology*. Tokyo: Kaitakusha.

Haraguchi, S. 1999. Accent. In *The Handbook of Japanese Lingusitics*, Natsuko Tujimura (ed.), Oxford: Blackwell, 1-30.

Hao, Y. 2008. Intonational interference on the perception and production of Mandarin Chinese tones by English-speaking second language learners. *Journal of the Acoustical Society of America,* 124, 2595 (A).

Hayes, B. 1995. *Metrical Stress Theory.* Chicago: The University of Chicago Press.

Hayes, B. 2008. *Introductory Phonology.* Oxford: Blackwell.

Howie, J. 1976. *Acoustical Studies of Mandarin Vowels and Tones* .Cambridge University Press, Cambridge.

Heuven V.J.J.P. van & Sluijter A.M.C. 1996. Notes on the phonetics of word prosody. In Goedemans R.. Hulst H. van der, Vische E. (ed.), *Stress patterns of the world, part 1: background*. The Hague: Holland Academic Graphic. 233-269.

Heuven, V.J .van. 1986. Perception of stress pattern and word recognition: Recognition of Dutch Words with incorrect stress position. *Journal of the Acoustical Society of America*, 78: s21.

Heuven, V. J. van.1988. Effects of stress and accent on the human recognition of word fragments in Spoken context: Gating and shadowing, *Proceedings of Speech 88, 7th FASE symposium*: 811–818. Edinburgh.

Hirayama, T. 1957. *Nihongo onchoo no kenkyuu*. Tokyo: Meiji Shoin.

Hyman, L.M. 2009. How (not) to do phonological typology: the case of pitch-accent. *Language Sciences* 31, 213-238.

Ishizawa, T. 2011. The effects of heavy syllable position and accentual pattern on the perception of Japanese special moras by L2 learners of Japanese. The Journal of Educational Research 8, 21-30

Johnson, K. & Babel, M. 2007. Perception of fricatives by Dutch and English speakers. *UC Berkeley Phonology Lab Annual Report 2007.*

Jongenburger,W. 1996.The role of lexical stress during spoken-word processing. Ph.D. dissertation, University of Leiden. The Hague: Holland Academic Graphics.

Jongenburger, W., & Heuven, V. J. van . 1995a.The role of linguistic stress in the time course of word recognition in stress-accent languages. *Proceedings of the 4th European Conference on Speech Communication and Technology*: 1695–1698). Madrid.

Jongenburger,W., & Heuven, V. J. van. 1995b. The role of lexical stress in the recognition of spoken words: Prelexical or postlexical? *Proceedings of the 13th International Congress of Phonetic Sciences* : pp.368–371. Stockholm.

Kager, R. 1989. *A Metrical Theory of Stress and Destressing in English and Dutch.* Dordrecht: Foris Publications Holland.

Kawahara, S. 2015. The phonology of Japanese accent. In *The Handbook of Japanese Phonetics and Phonology*. ed. by H. Kubozono. Mouton de Gruyter Series Handbooks of Japanese Language and Linguistics.

Kim, M.-R. & Duanmu, S. 1999. The interaction between consonant types and tone: a tonogenesis effect in Chonnam Korean. *Korean Journal 10*.

Kiriloff, C. 1969. On the auditory discrimination of tones in Mandarin, *Phonetica* 20: 63-67.

Kubozono, H. 1989. The mora and syllable structure in Japanese: Evidence from speech errors. *Language & Speech*, 32, 249-278.

Kubozono, H. 2008. Japanese accent. In *The Oxford Handbook of Japanese Linguistics*, S. Miyagawa & M. Saito, eds., Oxford: Oxford University Press, 165–191.

Labrune, L. 2012. *The phonology of Japanese*. Oxford: Oxford University Press.

Ladd, B. 1996. *Intonational Phonology*. Cambridge: CU Press.

Ladefoged, P & Maddieson, I. 1990. Vowels of the worlds' languages. *Journal of Phonetics* 18, 93-122.

Leather, J. 1990. Perceptual and productive learning of Chinese lexical tone by Dutch and English speakers. In Jonathan Leather & Allan James (Eds.), New Sounds 90: *Proceedings of the Amstesterdam*: University of Amsterdam.

Lehiste, I. (1970) *Suprasegmentals*. Cambridge, MA: MIT Press.

Lehiste, I. & Peterson, G.E. 1961. Some basic considerations in the analysis in the intonation. *Journal of the Acoustic Society of America,* 33: 419-423.

Lee, C. 2007. Does Horse Activate Mother? Processing Lexical Tone in Form Priming. *Language and Speech,* 50, 101-123.

Lee, C., Tao, L., & Bond, Z. S. 2008. Identification of acoustically modified Mandarin tones by native listeners. *Journal of Phonetics*, 36, 537-563.

Lee, L. & Nusbaum, H. C. 1993. Processing interactions between segmental and suprasegmental information in native speakers of English and Mandarin Chinese. *Perception and Psychophysics*, 53, 157-165.

Lee, Y., Vakock, D., & Wurm, L. 1996. Tone perception in Cantonese and Mandarin: A cross-linguistic comparison. *Journal of Psycholinguistic Research*, 125, 527–542.

McAllister, R., Flege, J. E., & Piske, T. 2002. The influence of L1 on the acquisition of Swedish quantity by native speakers of Spanish, English and Estonian. *Journal of Phonetics*, 30, 229-258.

Miyawaki, K., Strange, W., Verbrugge, R. R., Liberman, A. M., Jenkins, J. J., & Fujimura, O. 1975. An effect of linguistic experience: the discrimination of [r] and [l] by native speakers of Japanese and English. *Perception and Psychophysics*, 18, 331-340.

Nooteboom, S.G. 1972. Production and perception of vowel duration. Doctoral dissertation. Utrecht University.

Nooteboom, S.G. & Doodeman, G.J, N. 1985. Cues for lexical stress recognition of polysyllabic words, synthesized from diphones and presented in isolation. *In the proceedings of 109th Meeting of the ASA, Austin, TX*, 1985.

Otake, T, Hatano, G., Cutler, A. and J. Mehler. 1993. Mora or Syllable? Speech segmentation in Japanese. *Journal of Memory and Language* 32, 258-278.

Rohela-Madrazo, M. 2013. Perceptual Assimilation of Occluded Voiced Stops by Spanish Listeners. *Selected Proceedings of the 15th Hispanic Linguistics Symposium,* ed. Chad Howe et al., 140-156. Somerville, MA: Cascadilla Proceedings Project.

Pierrehumbert, J., & Beckman, M. (1988). Japanese tone structure. *Linguistic inquiry monograph 15*. Cambridge: MIT Press.

Quené, H. 1992. Durational cues for word segmentation in Dutch. *Journal of Phonetics*, 20 (3): 331-350.

Quené, H. & Koster, M. L. 1998. Metrical segmentation in Dutch: Vowel quality or stress? *Language and Speech*, 41, 185–202.

Repp, B. H. & Lin, H. 1990. Integration of segmental and tonal information in speech perception: A cross-linguistic study. *Journal of Phonetics*, 18, 481-495.

Rietveld, A. (1980) Word boundaries in the French language, *Language and Speech*, **23**, 289-296.

Schaefer, V. & Darcy, I. 2013. Linguistic prominence of pitch within the native language determines accuracy of tone processing. Selected Proceedings of the 2012 Second Language Research Forum, ed. Ryan T. Miller et al., 1-14. Somerville, MA: Cascadilla Proceedings Project.

Schaefer, V. & Darcy, I. 2014. Lexical function of pitch in the first language shapes cross-linguistic perception of Thai tones. *Laboratory Phonology 5 (4),* 489-522

Shibata, T. 1994. Gairaigo-ni okeru akusentokaku-no ichi [The location of the accent nucleus in Western words]. In *Kokugo Ronkyuu 4 Gendaigo–Hougen no Kenkyuu*, Kiyoji Sato, ed., Tokyo: Meiji Shoin, 388–418.

Sluijter, A.M.C. & Heuven V.J.J.P. van. 1995. Spectral balance as an acoustic correlate of linguistic stress. *Journal of Acoustical Society of America* 100: 2471-2485

Sluijter A.M.C. & van Heuven, V.J.J.P. 1996. Acoustic correlates of stress and accent in Dutch and American English. In NN NN (ed.) *Proceedings of the International Congress of Spoken Language Processing*. Philadelphia: NN.

Schwartz, B. 1993. On explicit and negative data effecting and affecting competence and linguistic behavior. *Studies in Second Language Acquisition* 15: 147-63.

Sheldon, A. & Strange, W. 1982. The acquisition of /r/ and /l/ by Japanese learners of English: Evidence that speech production can precede speech perception. *Applied Psycholinguistics*, 3, 243-261.

So, C. K. 2006. Perception of non-native tonal contrasts: Effects of native phonological and phonetic influences. In Paul Warren & Catherine I. Watson (Eds.), *Proceedings of the 11th Australian International Conference on Speech Science & Technology*. Auckland, New Zealand: University of Auckland.

So, C. K. 2010. Categorizing Mandarin tones in Japanese pitch-accent categories: The role of phonetic properties. Paper presented at *Interspeech 2010 Second language studies: Acquisition, learning, education and technology*. Waseda University, Tokyo, Japan.

So, C. K. & Stagray, J. R. & Downs, D. 1993. Differential sensitivity for frequency among speakers of a tone and a nontone language. *Journal of Chinese Linguistics*, 21, 143-163.

So, C.K & Best, C. T. 2010. Cross-language perception of non-native tonal contrasts: effects of native phonological and phonetic influences. Language and Speech, 53, 273-293.

Stagray, J. R. & Downs, D. 1993. Differential sensitivity for frequency among speakers of a tone and a nontone language. *Journal of Chinese Linguistics, 21*, 143-163.

Trommelen, M. & Zonneveld, W. 1999. Word stress in English and Dutch in *Word prosodic systems of the languages of Europe* ed. by H. van der Hulst, 132-3. Berlin: Mouton de Gruyter.

Tuc, H-D. 2003. *Vietnamese-English Bilingualism: Patterns of Code-switching*. London: Routledge-Curzon.

Utsugi, A., Koizumi, M., & Mazuka, R. 2012. The perception of non-native lexical pitch accent by speakers of “accentless” Japanese dialects. *In the proceedings of Speech Prosody 2010*, Illinois.

van Oostendorp, M. 2012. Quantity and the three-syllable window in Dutch word stress. *Language and Linguistic Compass* 6/6: 343-358.

van Selst, M. & Jolicoeur, P. 1994. A Solution to the Effect of Sample Size on Outlier Elimination, *The Quarterly Journal of Experimental Psychology*, 47A, 631-650.

Vance, Timothy J. (1995) Final accent vs. no accent: Utterance-final neutralization in Tokyo Japanese. *Journal of Phonetics* 23(4): 487–499.

Veenker, T.J.G. 2012. The Zep Experiment Control Application. Utrecht Institute of Linguistics OTS, University University.

Vũ, T. P. 1981. *The acoustic and perceptual nature of tone in Vietnamese*. Doctoral dissertation. Australia National University, Canberra.

Warner, N. 1997. Japanese final-accented and unaccented phrases. *Journal of Phonetics* 25: 43–60.

Wayland, R. P. & Guion, S. G. 2004. Training English and Chinese listeners to perceive Thai tones: a preliminary report. *Language Learning* 54:4, 681-712

Whalen, D.H. & Levitt, A.G. 1995. The universality intrinsic F0 of vowels. Journal of Phonetics 23, 349-366.

Wang, W. S-Y. 1976. Language change. In Stevan R. Hamad, Horst D. Steklis, & James Lancaster (Eds.), *Origins and Evolution of Language and Speech. Annals of the New York Academy of Sciences, 280* : 61-72.

Wang, Y., Allard, J. & Sereno, J. A. 2006. L2 acquisition and processing of Mandarin Chinese tones. In P. Li, E. Bates, L.H. Tan, & O. Tseng (Eds.), *The Handbook of East Asian Psycholinguistics* (pp. 250-256). Cambridge: Cambridge University Press.

Wang, Y., Spence, M. M., & Sereno, Joan A. 1999. Training American listeners to perceive Mandarin tones. *Journal of the Acoustical Society of America, 106*, 3649-3658.

Wu, X.H., & Lin, H. 2008. Perception of Mandarin Tones by Mandarin and English Listeners. *Journal of Chinese Language and Computing,* 18(4), 175-187.

Xu, Y, Gandour, J.T. & Francis, A.L. 2006. Effects of language experience and stimulus complexity on the categorical perception of pitch direction. *Journal of Acoustic Society of America* 120, 1063-107.4

Yamada, R., Kobayashi, N. & Tohkura, Y. 1996. Effect of word familiarity on non-native phoneme perception: Identification of English /r/, /l/ and /w/ by native speakers of Japanese. In A. James and J. Leather (eds.), *Second Language Speech*. New York: Mouton de Gruyter.

Appendix

Stimuli for Experiment 1 (ABX discrimination task)

Set 1

Contrast in two-syllable nonwords		
nono’ - nono	nono’ – no’no	no’no – nono
Contrasts in three-syllable nonwords		
nonono’ - nonono	nonono’ - no’nono	nonono’ – nono’no
nonono – no’nono	nonono – nono’no	no’nono – nono’no
Contrasts in four-syllable in four-syllable nonwords		
nononono’ – nononono	nononono’ – no’nonono	nononono’ – nono’nono
nononono’ –nonono’no	nononono – no’nonono	Nononono – nono’nono
nononono–nonono’no	no’nonono –nono’nono	no’nonono –nonono’no
nono’nono – nonono’no		

Set 2

Contrast in two-syllable nonwords		
yoyo’ - yoyo	/yoyo’/ –/yo’yo/	/yo’yo/ – /yoyo/
Contrasts in three-syllable nonwords		
yoyoyo’ - yoyoyo	yoyoyo’ - yo’yoyo	yoyoyo’ – yoyo’yo
yoyoyo – yo’yoyo	yoyoyo – yoyo’yo	yo’yoyo – yoyo’yo
Contrasts in four-syllable in four-syllable nonwords		
yoyoyono’ - yoyoyoyo	yoyoyono’ –yo’yoyoyo	yoyoyono’ – yoyo’yoyo
yoyoyono’ – yoyoyo’yo	yoyoyoyo – yo’yoyoyo	yoyoyoyo – yoyo’yoyo
yoyoyono– yoyoyo’yo	yo’yoyoyo –yoyo’yoyo	yo’yoyoyo – yoyono’yo
yoyo’yono – yoyono’yo		

Stimuli for Experiment 2 (Sequence recall task)

Positional Contrast	Token A	Token B
Initial accented vs. Final accented	/no’no/	/nono’/
Initial accented vs. Unaccented	/no’no/	/nono/
Final accented vs. Unaccented	/nono’/	/nono/

Two-word sequence: A-B, B-A, A-A, B-B, A-B, B-A, A-A, B-B

Three-word sequence: A-A-A, A-A-B, A-B-A, A-B-B, B-B-B, B-B-A, B-A-B, B-A-A.

Four-word sequence: A-A-B-A, A-A-B-B, A-B-A-A, A-B-B-A, B-B-A-B, B-B-A-A, B-A-B-B, B-A-A-B.