Research on the Processing of Nasal Sounds by Japanese Listeners*

Naoya Mori

要旨

本研究では日本語母語話者による英語音声の順行処理について実験を行い検証したものである。実験１として日本語母語話者が母語と同様の過程で英語の語末鼻音から情報を得て言語処理を促進しているのか検証したが、その証拠は得られなかった。実験１の再分析において語末音の長さが重要である事が示唆され、それに基づき実験２で検証を行った。結果英語の語末音が日本語の語末音と同等に長い時に言語処理が促進されている事を示す結果が得られ、聞き手の母語の知識を元に彼らは外国語音の言語処理を言語固有な方法で行っている事が明らかになった。

1. Introduction

The present paper reports the investigations as to how Japanese listeners use knowledge of their native language in the processing of English assimilated word-final nasal segments. Especially this paper focuses on facilitation effect caused in progressive processing of languages. To achieve this, the experiments employed phoneme monitoring tasks. The details of the experiments are described in Chapter 3 and 4.

Previous research has shown that syllable-final segments enable listeners to anticipate an upcoming segment (Gaskell & Marslen-Wilson, 1996; Gow, 2001; Lahiri & Marslen-Wilson, 1991; Mori, 2013; Otake et al., 1997). The listener picks up phonetic information, such as the place or manner of articulation, and predicts the upcoming segment based on the information in order to facilitate language processing. Since this procedure often involves assimilation across words, it is inferred that assimilation triggers of this anticipation.

The same procedure is similarly observed among Japanese listeners. According to Otake et al. (1996), Japanese listeners can anticipate which segment comes next based on the information of the preceding moraic nasals within a word. For example, a word *tombo*, 'dragonfly', contains moraic [m], and the [m] allows listeners to expect that a bilabial segment is coming. Similar processing is found when the moraic nasal and its subsequent segment are placed across words. For instance, a phrase like *gohon moeta* [gohom moεtɑ] 'five sticks burned' (Mori, 2013). Since Japanese nasal segments always undergo assimilation with their following segment, except for when the nasal segments are placed before pauses (Vance, 2008), anticipation by Japanese listener is sometimes referred to as an example of how assimilated segments work in the progressive processing.

It has been discussed how deeply progressive processing involves native language knowledge of listeners, and there are two views in language specificity of the processing. Those views are reviewed in chapter 2, and then language specificity of the processing by Japanese listeners are studied in the subsection. The following Chapter 3 examines the language specificity through an experiment.

In the process of investigation, it becomes clear that another factor drives language processing of Japanese listeners. Discussion and an investigation are in section 3.5 and later.

2. Theoretical backgrounds

2.1 Perceptually universal or Language specific? Two views in language processing

Although there have been numbers of successful research on language processing in listeners' native language, investigations involving listeners' foreign language are not conclusive. The original attempt of employing participants' foreign languages in past investigations was to prove that a theory of the processing mechanism is universal to any languages. The attempt, however, narrowly succeeded. The results of the previous research were thus split into two: perceptually universal and language specific. Indeed, research which states that the mechanism of language processing is perceptually universal suggests that an assimilated segment informs its phonetic detail to listeners. Research such as Mitterer et al. (2006), Tsukada (2005) and Tsukada et al. (2007) had provided evidence of this approach, but most of them report successes of listeners' exploitation of their foreign assimilated segments. Few researchers modeled a mechanism, but Gow (2003) proposed one called the Feature Cue Parsing. A crucial point of the mechanism is that assimilation is a gradient; a segment undergoes incomplete assimilation (Zsiga, 1995; Stephenson et al., 2002). Therefore, the segment will contain its underlying information and its following segmental information. For example, in a phrase *cone belt*, the *cone* assimilates to the first segment of its following word and is pronounced as [kóʊm]. Since the assimilation is half complete, the sound of the final segment shifts from its unmodified [n] to assimilated [m]. The Feature Cue Parsing expects that the phonetic transition is perceptible. All that listeners are required to do is to listen, and the segment tells its origin, its alternation, and what is coming next. When listeners hear bilabial sound in the final part of the transition, they are informed that a labial segment is coming next, a type of assimilation that is often observed in English. Therefore, Gow (2003) uses English participants and English stimuli in his experiment. This gradience with the assimilated segment, however, can be a flaw when applying the mechanism of Gow (2003). As the mechanism is based on the gradience, it cannot be applied to languages in which the assimilation includes complete alternation (Darcy et al., 2009). In the current research, however, the model is useful because stimuli that are used in the current experiment were prepared in English whose assimilation is incomplete.

However, while noting this, it is also claimed that the processing is language specific when listeners behave in the same way to any languages as in their native language. Support for this claim comes from the experiment performed by Darcy et al. (2009). Participants were native English speakers and native French speakers. They were required to hear a sequence of English and French stimuli which were conditioned in three patterns: unmodified, French voicing assimilation, and English place assimilation. They were then instructed to hear a pair of participants' native language stimuli that consisted of unmodified and either French voicing assimilated or English place assimilated and were then asked to judge whether the second phrase sounded like the first one. The result showed that French participants found the stimuli identical when the second one was conditioned with French assimilation, and English participants found the stimuli identical when the second one was conditioned with English assimilation. This result indicated that listeners preferred to use their native knowledge for language processing. In other words, the English participants felt natural when listening to phrases modified with English place assimilation such as *wet pants* [wɛp pænts] but not to phrases irregularly modified with French voicing assimilation such as *black glove* *[blæg glʌv]. The result of Darcy et al. (2009) indicated that listeners could not use the linguistic knowledge that is not included in their native language. They were, therefore, unable to process assimilation of foreign languages.

Other avenues of support for the language-specific theory include research conducted by Mori (2014) and Mori (2015). These studies tried to figure out whether the processing strategy of Japanese listeners' was perceptually universal or language-specific, mainly through an experiment using a cross-modal lexical judgment task with English stimuli that contained ambiguity caused by place assimilation. Mori (2015) expected that Japanese listeners' strategy should be language-specific since Japanese moraic consonants always assimilate to their subsequent segment spontaneously, and Japanese word-final consonants were limited to only /ɴ/ and special cases of /**Q**/ such as *だめっ！* /**dameQ**/ 'No!', unless phonetic alternation occurs. The prediction was proven to be on the right track, as it indicated that Japanese participants compensated English assimilation of word-final nasals in the Japanese way. In other words, they interpreted every English word-final nasal as alveolar nasal /**n**/ so that [**gέɪm**] could be interpreted as *gain*, even when the word was *game*.

While participants behaved as expected to nasal sounds, they did not act in the same way to stop consonants. In the experiment of Mori (2015), participants could not recover the underlying forms of assimilated stops. Instead, they appeared to have interpreted the word-final stop to have the same place of articulation (POA) as its following segment. For example, in a phrase like *cap traded* [**kǽp tréɪdɪd**], [**kæp**] was mistakenly interpreted as /**kæt**/. Mori (2015) names the unexpected processing of word-final stop consonants as "filling-up behavior" . He hypothesizes that Japanese listeners transfer phonetic information, such as POA, from a post-assimilated initial consonant to its previous assimilated word-final stop. This phenomenon might have something to do with a structural difference between English and Japanese, especially with their patterns of segmentation. English is known to be a stress-timed language in which listeners put a boundary for segmentation before stressed syllable (Cutler & Butterfield, 1992) while Japanese is a mora-timed language in which listeners find a boundary between every mora (Otake et al., 1993). This difference of segmentation suggests that a word consisted of CVCCV (C stands for consonant, and V for vowel) is segmented as CVC.CV among English listeners, meanwhile Japanese listeners segment it as CV.C.CV. As mentioned earlier, a potential syllable-final stop consonant in Japanese is limited to /**Q**/, which always assimilates completely to its succeeding segment unless it is prior to a pause. It is quite natural for Japanese listeners to perceive the coda C in the first syllable as an assimilated /**Q**/. In addition, it could also have something to do with a phenomenon that word-final consonants become unreleased in common circumstances, especially when two stops come in the sequence (Henderson & Repp, 1982). Although this behavior has not been discussed sufficiently, I will not pursue the mechanism of this behavior in this paper. Rather, it can be stated that aside from word-final stops, Mori (2015) suggests that the Japanese strategy of the disambiguating process is language specific.

2.2 Language processing by Japanese listeners: Prediction

Disambiguation and anticipation are both sides of the same coin. Disambiguation and anticipation involve assimilation, and, in the meantime, assimilation contributes to the process of disambiguation and anticipation. It is very natural that listeners use assimilation in their native language for foreign language processing since cross-language transfer of native language structure has been well investigated (Durgunoğlu et al., 1993; Gass & Selinker, 1983; Kellerman & Sharwood, 1986; Odlin, 1989), but how do they apply native language assimilation to non-native language processing? One answer to that question comes from a model called Perceptual Assimilation Model (PAM), proposed by Best & Tylor (2011). In PAM, it is considered that every speaker has a phonological inventory of his/her native language. When the speaker hears a foreign sound for him/her, he/she checks phonetic details of the sound and categorizes the sound to perceive it as a native phoneme whose feature is the closest to the input. In other words, PAM is a model to explain how a human hears foreign sounds and interprets them. According to the model, foreign sounds are interpreted in four ways: single category, two categories, category goodness, and uncategorized. The single category indicates two sounds are interpreted as one phoneme in a listener's mind. For example, Japanese listeners do not distinguish [l] and [r], and both segments are recognized as a phoneme /ɾ/. The two categories express that the two sounds are recognized as two individual sounds. For example, [t] and [p] are recognized as /t/ and /p/. In category goodness, two different sounds are heard as a phoneme, but a listener distinguishes them. The perfect example of this interpretation is place assimilation. For example, a word *lean* [li:n] potentially gets assimilated when it is followed by another word such as *bacon* and its pronunciation will change to [li:m]. In this case, both word-final segments, [n] and [m], are perceived as /n/. At the same time, as stated earlier, the assimilated [li:m] provokes anticipation that an upcoming segment is bilabial. It indicates that listeners hear [li:m] as /li:n/, but still they can sense bilabial feature in the assimilated word-final segment and exploit it for language processing. The last way, uncategorized, indicates that there are no similarities between the inputs and categories in a listener's mind.

PAM predicts that Japanese listeners can process English assimilated segments and use them to anticipate upcoming segments. This assumption stems from the nature of the Japanese final consonants; they are individual phonemes /ɴ/ and /Q/, and always assimilate completely to the following segment. According to PAM, Japanese listeners will categorize the English final consonant which has undergone assimilation into either within a single category or category goodness. This assumption is based on the fact that Japanese syllable-final consonants are derived from either /ɴ/ or /Q/, a key point as to whether Japanese listeners distinguish each assimilated form of the syllable-final consonant or not. Indeed, there is a good reason to believe they can distinguish each realization. As Mori (2013) reported, Japanese listeners anticipate the following context based on moraic nasals in Japanese. For example, the word-final moraic nasal in *gohon moeta* [gohom moɛtɑ] 'five sticks burned' and the one in *gohon katta* [gohoŋ kɑtːɑ] 'bought five sticks' are the realization of the phoneme /ɴ/. Those moraic nasals, however, make Japanese listeners anticipate different segments. The former, *gohon* [gohom], provokes anticipation of an upcoming bilabial, and the latter, *gohon* [gohoŋ], provokes anticipation of an upcoming velar. The observation of Mori (2013) suggests that Japanese listeners hear the difference between [gohom] and [gohoŋ]. Since the assimilation rule in Japanese shares a lot with that of English ones, for instance, as both Japanese moraic nasal /ɴ/ and English syllable-final nasal /n/ assimilate to the following bilabial and result in a bilabial nasal [m], it is very reasonable to believe that English assimilated nasals do not inhibit anticipation by Japanese listeners.

On the other hand, however, Japanese listeners' reactions to English syllable-final stops are not predictable. Logically speaking, English syllable-final stops should be categorized as /Q/ by Japanese speakers, according to PAM. This logic, however, had already been denied in Mori (2015), as participants in his experiment showed the application of filling-up behavior in which listeners tolerate features of assimilated stop segments and transfer upcoming segments' features instead. As the earlier example showed, they could not recover the underlying phoneme of assimilated stops and interpreted stimuli as these were heard. The inconsistency between the PAM theory and the empirical result thus makes it very hard to expect what happens in the current investigation. As such, I adopted the idea propagated by Mori (2015) to predict the result. The idea predicts that Japanese listeners cannot make anticipations with word-final stops. As was stated, word-final consonants are unreleased in most cases. Thus, when stop sounds become unreleased, they remain closed and consequently make a soundless "gap" in a speech. For example, a phrase *cap bought* [kæp̚ bɔːt] can sound like [kæp̥ bɔːt]. It is possible that Japanese listeners cannot obtain any exploitable information due to the occurrence of this "gap" . Indeed, Tsukada (2005) has reported that Japanese listeners could not hear unreleased Thai word-final stops although they could discriminate released Australian English word-final stops. The subsequent study by Tsukada et al. (2007) reported a similar result, yet neither of them tried to explain why.

So far, the prediction is that Japanese listeners can exploit information in English assimilated final nasal consonants and use it to anticipate upcoming segments but cannot access to assimilated final stop to gain following phonetic information. This partial use of word-final consonants characterizes Japanese listeners' language processing manner as language specific, for the predicted reaction to English word-final consonants by Japanese does not identify with that of native English listeners' as in Gaskell & Marslen-Wilson (1996) and Gow (2001). In the following sections, I will report the findings of my phoneme monitoring task experiments and discuss the capabilities of Japanese listeners' in the use of English phonetic information.

3. Experiment 1

3.1 Method

The purpose of this experiment was to investigate whether Japanese listeners could pick up the phonetic information from the English unreleased word-final assimilated segment and whether they used it for the anticipation of the post-assimilation context. Thus, a phoneme monitoring task was used for this experiment. The phoneme monitoring task required participants to listen to audio stimuli streamed from the headphone, which was connected to a personal computer. They were then instructed to press a button when they found a target phoneme displayed on the computer, and the time duration from the onset of the target to the participants' reaction was recorded as the reaction time. With the phoneme monitoring task, we would be able to investigate how conditioned stimuli influenced the participants' online processing. When participants hear an assimilated pre-target sound, it is said that they pick up the phonetic information of the pre-target sound and anticipate the upcoming phoneme, which would be an initial segment of the target in this case. Since the anticipation depends on the information of the pre-target, we can modify the pre-target to mislead the participants' expectation. As it is assumed that reaction time to misleading stimuli is conditionally neutral, we can collect reaction time that is free from the influence of anticipation. We can statistically compare the reaction time to conditioned stimuli with that to misleading stimuli. When we find the reaction time to conditioned stimuli are significantly shorter than the reaction times to misleading stimuli, we can say that there is facilitation effect caused by anticipation. Otherwise, it is assumed that participants failed to use phonetic information, and consequently, no anticipation occurred.

In Experiment 1, participants saw an English letter which was either t, d, p, or b on the computer screen, and they were required to hear an English sentence such as *These words on the wall warn people in the falling tunnel*. In this case, the letter p was presented on the screen and participants were required to press a button when they heard *people* in the sentence.

3.2 Stimuli

The English words whose final consonant was either the alveolar nasal (e.g., *warn* [wɔːn]), the alveolar stop (e.g., *cat* [kæt]), the bilabial nasal (e.g., *warm* [wɔːm]), or the bilabial stop (e.g., *cap* [kæp]) were chosen as pre-targets. They were paired with the words whose initial sounds were either alveolar (e.g., *travelers / discounted*) or bilabial (e.g., *people / picked*) to form phrases. Those paired words were chosen to be targets in the experiment. The phrases were then embedded into sentences (e.g., *These words on the wall warn people in the falling tunnel. / He bought a cat picked from the store.*). An American male read these sentences, and his reading was recorded with a MacBook Pro (13-inch, early 2011), Roland QUAD-CAPTURE UA-55, and audio-technica microphone AT4040. The recording was operated on an audio software Audacity and digitized at 24bit/48Hz. After the recording, the recorded sentences were cross-spliced between the pre-target and target. Specifically, each sentence had its minimal pair group. For example, a sentence such as *These words on the wall warn people in the falling tunnel* had its grouped sentences *These words on the wall warm people in the* *falling tunnel* and *These words on the wall warn travelers in the falling tunnel*. The cross-splicing process switches the former parts before targets among grouped sentences and makes a sentence where the place of articulation of its pre-target and its target does not match phonetically. For instance, in a cross-spliced sentence *These words on the wall warm travelers in the falling tunnel*, the final [m] in *warm* and the initial [t] in *travelers* did not match in their POA, as it was assumed that it would not provoke anticipation. This cross-splicing process was operated on Praat and resulted in two levels of conditions: 2 POA-match conditions and 2 POA conditions. Table 1 summarizes these experimental conditions. After all, there were 12 pre-targets, and each one had four variations. Therefore 48 stimuli were used in Experiment 1.

Table 1. *Experimental Conditions in Experiment 1*

POA-match	POA	
	nasal	stop
matched	warn travelers / warm people	cat traded / cap purchased
mismatch	warn people / warm travelers	cat purchased / cap traded

3.3 Participants

The participants were 19 undergraduate students. Among them, 6 were male, and 13 were female. The average age was 20.3.

3.4 Analysis

The statistical analysis was conducted to examine if there were differences between the reaction time to POA conditions and that of POA-match conditions, and interaction between these conditions through a mixed linear model in R program. The result is shown in Table 2.

The mean reaction time to the nasal POA condition was 440.667ms (SD = 154.45), and that of the stop POA condition was 452.875ms (SD = 154.509). There was no significant difference between POA conditions (t = .982, df (40.478), $p > 0.1$). The mean reaction time to matched was 437.88ms (SD = 150.876) and that of mismatch was 459.089ms (SD = 158.011). There was also not significant difference between POA-match conditions (t = 1.24, df (40.395), $p > 0.1$). Interaction was not observed between conditions (t = .116, df (40.304) $p > 0.1$). This analysis presented a marginally large-

Table 2. *Fixed Effects of the Final Linear Mixed Regression Model Performed in Experiment 1*

conditions	Expression in the formula below	Estimation	*SE*	*t*	*p*
Intercept		445.815	15.748	28.309	.000***
POA	f1c	13.061	13.299	.982	.332
POA-match	f2c	16.345	13.179	1.24	.222
interaction	fc1*fc2	3.049	26.345	.116	.908

Note. This result is obtained through the following formula: glmer (RT ~ f1c * f2c + (1 + f1c + f2c | participants) + (1 + f1c + f2c | trial))

sized effect (R^2 = 0.252).

3.5 Discussion

The result showed that there were no facilitation effects in any stimuli condition. The result indicated that Japanese listeners did not perceive any information from the English assimilated segments, and as such, no anticipation was provoked. This result was thus inconsistent with the prediction. Although a similar experiment that was conducted in Japanese found the facilitation effect in nasal stimuli (Otake et al., 1996), my experiment did not find the facilitation effect even in the nasal stimuli. Since the nasal assimilation in English is almost the same as that in Japanese, except for its underlying form (the underlying form is always /ɴ/ in Japanese), this result appears strange. If the current situation is not a result of the design flaw, then there must be a missing factor that would be crucial to provoking anticipation beside assimilation of a syllable-final segment.

In order to seek out the factor, I studied some cross-language comparison researches and found some clues. In Hamann and Sennema (2005), the temporal duration of labiodental in German and Dutch was compared. In their research, German labiodental /v/ was about 100 ms long and /f/ was about 170ms long, while the Dutch /ʋ/ was about 100ms long, /v/ was about 140ms long, and /f/ was about 170ms long. In the subsequent study, Hamann (2009) pointed out that the German learners of Dutch had their own rule of categorizing Dutch /v/ until they could correctly categorize it since its auditory value was so different from German /v/. These reports indicated that foreign language sounds were not always heard as they are, and sometimes auditory values, like duration, influenced the listeners' categorization of foreign sounds.

Another clue was found in the investigation by Shoemaker (2014), which studied how French listeners clarified ambiguous phrases. For example, a phrase [œ̃.nɛʁ] can be interpreted in two ways: un nerf 'a nerve' or un air 'an air'. In the research, she modified the length of [n] in the phrases to make the stimuli. In the experiment, two stimuli were presented within a designed task to participants, and they were required to judge whether those two stimuli were identical or not. The result indicated that French listeners exploited the duration of [n] to disambiguate phrases.

A similar study was conducted in Fujisaki & Sugitou (1977), which aimed to find out durational parameters for Japanese listeners to distinguish *ama* [ɑmɑ] 'nun' from *amma* [ɑmːɑ] 'massage', for example. Logically speaking, since *amma* is a three-mora word meanwhile *ama* is a two-mora word, it is natural to assume that the inter-nasal part of *amma* is longer than that of *ama*. The reason why *amma* gets longer is that it has moraic nasal /ɴ/ in the middle, but it completely assimilates and connects to the following mora-initial /m/ to be impossible to detect a boundary between them. Fujisaki & Sugitou (1977) revealed that the parameter existed when between-vowel nasal duration was 152ms in a short sentence. When the duration was longer than 152ms, listeners perceived the word as *amma* otherwise *ama*. The research of Fujisaki & Sugitou (1977) indicated that Japanese listeners required specific duration to recognize moraic nasal, whose result was identical with Shoemaker (2014). They are convincing that durational cue has something to do with the results in Experiment1.

Furthermore, Sato (1993) reported three language comparisons among Japanese, English, and Korean. By comparing syllable-initial and syllable-final (within-mora and moraic, in the Japanese case) consonants and calculating the durational ratio, she found that the durational ratio of Japanese initial nasals and final nasals was about 1:2 while the English counterpart was about 1:1.3. Japanese final nasals were about 1.54 times longer than English final nasals. In other words, it is possible that English final consonants were too short for Japanese listeners to perceive phonetic information accurately. This direct clue made it confident that the duration was the missing factor in the current experiment.

3.6 Reanalysis

Since I learned that duration was a crucial factor for anticipation, I performed a statistical reanalysis and setting the duration of the pre-target final segments as a variable. The duration of the pre-target final segment of each stimulus was measured, and the stimulus was labeled 'long' when they were longer than the average value of the duration of pre-target finals; otherwise, they were labeled 'short'. Reaction times to each condition were compared through the t-test, and the result of the reanalysis is shown in Table 4.

Table 3. *The Result of the Statistical Reanalysis of Experiment 1*

	M (ms)	*SD*	*r*	*p*
Short	449.488	154.225	.13	.001
Long	409.016			

The result showed that the reaction time to the longer duration stimuli was significantly shorter than that of the shorter duration stimuli. ($t = 3.295$, df (588), $p < 0.01$, $r = .13$) It also indicated that Japanese listeners could use phonetic information when syllable-final segments had a longer duration.

3.7 New hypothesis

The reanalysis brought a new perspective to the study. Although exploitation of the segmental duration in the perception of syllable-final or moraic nasal was discussed in Shoemaker (2014) and Fujisaki & Sugitou (1977), there had been few studies that pointed out the Japanese listeners' use of duration in progressive processing. The current reanalysis thereby serves as preliminary evidence that duration is a factor that is crucial to triggering anticipation for Japanese listeners. This result leads us to a new hypothesis, namely that Japanese listeners need the specific duration of syllable-final nasal sounds to perceive them properly. For instance, the Japanese moraic nasal is 1.54 times longer than English syllable-final one, according to the observation by Sato (1993). In other words, English syllable-final nasal is possibly too short for Japanese listeners to perceive property, and they need longer word-final durations than English natural ones when they try to exploit phonetic information within the word-final segment.

Unfortunately, the result of the reanalysis cannot fully support the new hypothesis because it comes from the experiment that was not designed to investigate the effect of segmental duration on anticipation. While it can be certain that the result is a precursor to something of significance, it is not conclusive enough until the result can be recreated in another experiment. In order to examine the hypothesis, the second experiment was carried out. In the following chapter, the detail of the second experiment will be described.

4. Experiment 2

The second experiment was designed to examine the influence of the word-final segmental duration on the listeners' progressive processing. Therefore, the duration was manipulated to condition experimental stimuli in experiment 2. The task employed in Experiment 2 was the phoneme monitoring task, which is similar to that of experiment 1.

4.1 Stimuli

As pre-targets, English words whose final segments were alveolar nasal /n/, *son*, for example, were chosen. In experiment 2, I decided to use only nasal-final words to obtain an evidential effect of segmental duration. Although syllable-final stops had been used in my past experiments, the results were generally problematic and unexpected. Based on the discussion in Mori (2015) and the current result in experiment 1, it is safe not to use syllable-final stops. Exploitation of syllable-final stops should be discussed at another time, including an exposition of filling up behavior. English verbs whose final sound was /t/ as in *tear*, /d/ as in *dream*, /p/ as in *polish*, and /b/ as in *believe* were also selected to be targets in this experiment. Those pre-targets and targets made phrases as *son believes*, and then the phrases were embedded into a sentence like *I heard my son believes in Santa Clause*. An American male read these sentences, and his reading was recorded with a MacBook Pro (13-inch, early 2011), Roland QUAD-CAPTURE UA-55, and audio-technica microphone AT4040. The recording was operated on the Audacity program and digitized at 24bit/48Hz.

After the recording, the final /n/ in pre-targets were edited, and two patterns of experimental conditions were created: the natural pre-target type (natural, hereafter) and the slowed pre-target type (slowed, hereafter). The natural condition was the stimuli which were not modified at all. Therefore, this type of condition was prepared as a control condition. The slowed condition was synthesized in the Audacity app, in which the final nasal sounds were slowed down about 154% of the original length to be twice as long as an initial nasal. The slowing down rate was calculated based on the observation of Sato (1993), who claimed that the Japanese moraic nasal was twice as long as mora initial nasal, whereas English syllable-final ones were 1.3 times longer than English syllable-initial ones. This modification was applied from the onset of the nasal to the burst of the initial consonant of targets. This modification might not be conventional way, but I operated it for the matters of uniformity since some recorded sentences had no closure parts.

Each type of conditions was packaged separately: natural package and slowed package. A package included 40 sentences; therefore, 80 sentences were made in total.

4.2 Participants

Participants were 40 undergrad students. Among them, 20 were male, and 20 were female. The average age was 19.8.

4.3 Analysis

The reaction time from the onset of the target to a participant's pressing button was used in the analysis. The statistical analysis was conducted to examine if there was a difference between the reaction time to the natural stimuli and that of slowed stimuli, through the use of a mixed linear model in R program. The result is shown in Table 4.

Table 4. *Fixed Effects of the Final Linear Mixed Regression Model of RTs in Experiment 2*

conditions	Expression in the formula below	Estimation	SE	*t*	*p*
Intercept		590.6	20.554	28.734	.000***
Pre-target final	f1c	−98.55	32.417	−3.04	.003 **
Target initial	f2c	50.892	34.628	1.47	.151
interaction	fc1*fc2	−53.728	47.763	−1.2	.236

Note. This result is obtained through the following formula: glmer (fixedrt ∼ f1c * f2c + (1 + f1c + f2c | participants) + (1 + f1c + f2c | trial) + pretargetfamc + targetfamc + targetlocc)

The mean reaction time to the slowed condition was 543.5ms (SD = 264.503), and that of the natural condition was found to be 653.8ms (SD = 269.231). The former was significantly shorter than that of the latter (t = −3.04, df (49.503) p = 0.03). No significant difference was found between target initial type (t = 1.47, df (32.618) p > .1), and interaction was also not observed (t = 1.2, df (47.604) p > .1). The test presented a large-sized effect (R^2 = 0.345). 92ms of the difference of mean reaction time between packages indicated that a factor in slowed pre-targets facilitated listeners' speech perception, in this case, that was slowed word-final nasal in pre-targets. And thus, an induction that slowed pre-targets provoked anticipation was made since anticipation was a cause for facilitation effect in progressive processing. This result also indicated that the original length of English final nasal segments was too short to provoke anticipation. Therefore, it is considered that participants failed to perceive phonetic information when they heard natural stimuli.

4.4. Discussion

The result of experiment 2 shows that there is a significant difference between the responses to the natural condition and the slowed condition, and the reaction to the slowed condition was shorter than its opponent. This result can thus be used as evidence for the hypothesis that slowed final nasal provokes anticipation and subsequently facilitates progressive processing. In other words, Japanese listeners require a foreign word-final segment to be as long as Japanese ones when they try to exploit its phonetic information for anticipation. This result is consistent with the claim that Japanese listeners need a specific duration of the final segment to perceive it accurately. This result can also be a support to Darcy et al. (2009), which proposes that a listener's speech processing is language specific. Since listeners use their native knowledge of word-final duration, a procedure of the language processing is driven by language-specific factors.

Knowing that the duration is a vital factor in Japanese listeners' language processing, the next goal of the study is thus to investigate what is the role of duration in language processing. It has been considered that the assimilated segment contains following contextual information to a greater or lesser extent and therefore assimilation triggers anticipation. As such, it is assumed that the assimilated segment spontaneously provides clues of an upcoming segment to listeners. This assumption, however, was not entirely right. In Experiment 1, English assimilation did not supply any information to

Japanese listeners who could anticipate post-assimilation context in their native language. The reanalysis of Experiment 1 and the result of Experiment 2 demonstrated that the duration of the assimilated segment had something to do with the facilitation effect caused in the progressive processing. Since facilitation is an effect caused by anticipation, and the anticipation bases itself on the knowledge of assimilation, a hypothetical function of the duration in foreign language progressive processing is a criterion for screening out segments that are too short. It checks if a word-final segment is long enough and similar to that of listeners' native language and converts it to listeners' native counterpart. Once it is converted, the rest is the same as it has been discussed; listeners pick up phonetic information from the word-final assimilated segment and use it to accelerate processing. This new hypothetical mechanism, however, still need some supports before modeling. The current experiments investigated only assimilation-viable stimuli because it aimed to examine the effect of segmental duration. A further investigation involves assimilation-unviable stimuli is ongoing.

5. Conclusion

I have carried out two experiments and examined what triggers anticipation of post-assimilation phonetic context. Experiment 1 employed conditions based on the conventional way, but it turned out that it was not effective in the investigation of foreign language processing of Japanese listeners. Through a reanalysis, it was suggested that the duration of an assimilated segment was a factor in progressive processing. Experiment 2 thus examined the suggestion and found out the duration was crucial to language processing among Japanese listeners.

The current report was not sufficient to hypothesize the mechanism of Japanese listeners', but it became clear that the duration was the vital factor in their language processing. It is expected that the ongoing investigation unveils the function of the duration in the processing.

Notes

*I would like to thank Professor Satoru Nakai and the anonymous reviewers for the help during preparation of the paper. I also thank all the participants in the current experiments. All errors are mine.

References

Audacity Team. (2017). Audacity(R) : Free Audio Editor and Recorder [Computer application]. Version 2.2.1 retrieved 20 December 2017 from https://audacityteam.org/

Best, C. T., & Tyler, M. D. (2011). Nonnative and second-language speech perception: Commonalities and complementarities. In M.J. Munro & O.-S. Bohn (eds) *Second language speech learning: The role of language experience in speech perception and production*. Amsterdam: John Benjamins. (Accepted 16 Jan 2006).

Boersma, P. & Weenink, D. (2018). Praat: doing phonetics by computer [Computer program]. Version 6.0.43, retrieved 8 September 2018 from http://www.praat.org/

Cutler, A. & Butterfield, S. (1992) Rhythmic cues to speech segmentation: Evidence from juncture misperception. *Journal of Memory and Language*, 31, 218 - 236

Darcy, I., Ramus, F., Christophe, A., Kinxler, K., & Dupoux, E. (2009). Phonological knowledge in compensation for native and non-native assimilation. *Variation and gradience in phonetics and phonology*, 14, 265

Durgunoğlu, A. Y., Nagy, W. E., & Hancin-Bhatt, B. J. (1993). Cross-language transfer of phonological awareness. *Journal of educational psychology, 85*(3), 453.

Fujisaki, H., & Sugitou, M. (1977). Onsei no butsuri-teki tokusei [Physical characteristics of speech], *Iwanami kouza Nihongo 5 On-in* [*Iwanami lectures about Japanese linguistics, volume 5, Phonology*], 63-98, Iwanami Press.

Gass, S., & Selinker, L. (1983). *Language transfer in language learning*. Rowley, MA: Newbury House.

Gaskell, M.G., & Marslen-Wilson, W.D. (1998). Mechanisms of phonological inference in speech perception. Journal of Experimental Psychology: Human Perception and Performance, 24, 380–396.

Gow, D. W. (2001). "Assimilation and anticipation in continuous spoken word recognition," Journal of Memory and Language. 45, 133–159.

Gow, D. W. (2003). "Feature parsing: Feature cue mapping in spoken word recognition," Perception & Psychophysics 65, 575–590.

Hamann, S. & Sennema, A. (2005). Acoustic differences between German and Dutch labiodentals. ZAS Papers in Linguistics 42: 33-41.

Hamann, S. (2009). Variation in the perception of an L2 contrast: A combined phonetic and phonological account. *Phonology and Phonetics*, 71-98.

Henderson, J. B., & Repp, B. H. (1982). Is a stop consonant released when followed by another stop consonant?. *Phonetica, 39*(2-3), 71-82.

Kellerman, E., & Sharwood Smith, M. (1986). *Crosslinguistic influence in second-language acquisition*. Elmsford, NY: Pergamon Press.

Lahiri, A., & Marslen-Wilson, W.D. (1991). The mental representation of lexical form: A phonological approach to the recognition lexicon. Cognition, 38, 245–294.

Mitterer, H., Csépe, V., Honbolygo, F., & Blomert, L. (2006). The recognition of assimilated word forms does not depend on specific language experience. Cognitive Science, 30, 451–479.

Mori, N. (2013). Regressive and Progressive Effect of Assimilation in Speech Perception by Japanese. Doshisha University, submitted to a partial requirement for master degree

Mori, N. (2014). Nihon-jin eigo-gakushusha ni yoru eigo onsei-tango no gyakkou-shori [Japanese listeners' compensation on English assimilated words]. *KLS 38: Proceedings of the thirty-eighth annual meeting of the Kansai Linguistic Society*. 241-252.

Mori, N. (2015). Japanese Listeners' Compensation and the Use of Contextual Information at the Segmental Level. *Phonological Studies*, (18), 59-66.

Odlin, T. (1989). *Language transfer*. Cambridge, England: Cambridge University Press.

Otake, T., Hatano, G., Cutler, A., & Mehler, J. (1993). Mora or syllable? Speech segmentation in Japanese. *Journal of memory and language, 32*(2), 258-278.

Otake, T., Yoneyama, K., Cutler, A., & van der Lugt, A. (1996). The representation of Japanese moraic nasals. *Journal of the Acoustical Society of America*, 100, 3831-3842.

Sato, Y. (1993). The durations of syllable-final nasals and the mora hypothesis in Japanese. *Phonetica, 50*(1), 44-67.

Shoemaker, E. (2014). Durational cues to word recognition in spoken French. *Applied Psycholinguistics, 35*, 243-273.

Stephenson, L. & Harrington, J. (2002). Assimilation of place of articulation: Evidence from English and Japanese. *Proceedings of the 9th Australian International Conference on Speech Science & Technology*, 592-597. Melbourne

Tsukada, K. (2005). Cross-language speech perception of final stops by Australian-English, Japanese and Thai listeners. *Proceedings of ISCA Workshop on Plasticity* *in Speech Perception* (*PSP2005*), London. 244–247.

Tsukada, K., Ishihara, S., Nguyen, T. T., & Roengpitya, R. (2007). Cross-language perception of word-final stops: A comparison of Cantonese, Japanese, Korean, and Vietnamese listeners. *Acoustical Society of America Journal, 121*, 3074-3074.

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

Zsiga, E. C. (1995). An acoustic and electropalatographic study of lexical and post lexical palatalization in American English. In (B. Connell & A. Arvaniti eds.) Phonology and Phonetic Evidence: Papers in Laboratory Phonology IV, 282-302. Cambridge, Cambridge University Press.