Korean Journal of English Language and Linguistics, Vol 23, February 2023, pp. 175-191
DOI: 10.15738/kjell.23..202302.175

KOREAN JOURNAL OF ENGLISH LANGUAGE AND LINGUISTICS

ISSN: 1598-1398 / e-ISSN 2586-7474

http://journal.kasell.or.kr

Word Recognition in English Place Assimilation by L1 and L2 Listeners: An Eye Tracking Study*

Eunkyung Sung(Cyber Hankuk University of Foreign Studies**) Sehoon Jung**(Kyungsung University**)**
Sunhee Lee(Cyber Hankuk University of Foreign Studies)

Received: January 16, 2023
Revised: February 10, 2023
Accepted: February 18, 2023

Eunkyung Sung (1st author)
Professor, Dept. of English, Cyber Hankuk University
E-mail: eks@cufs.ac.kr

Sehoon Jung (corresponding author)
Lecturer, Dept. of English, Kyungsung University
E-mail: sejung@ks.ac.kr

Sunhee Lee (co-author)
Associate Professor, Dept. of Chinese, Cyber Hankuk University
E-mail: lishanxi@cufs.ac.kr

* This paper was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2019S1A5A2A01047388).

ABSTRACT

Sung, Eunkyung, Sehoon Jung and Sunhee Lee. 2023. Word recognition in English place assimilation by L1 and L2 listeners: An eye tracking study. *Korean Journal of English Language and Linguistics* 23, 175-191.

This study explores the dynamics of lexical activation by comparing the time course of word recognition between assimilated forms (e.g., *ca[^{t}p]* in *cat box*) and noncoronal forms (e.g., *ca[p]* in *cap box*). Using the Visual World Paradigm, an eye-tracking method, the main goal was to investigate how gradient modification in place assimilation context influences L1 and L2 listeners' real time word recognition in English. Twenty native Korean learners of English, as well as fourteen native English listeners took part in the listening task integrated into the eye-tracking experiment. The participants were given aural input in the form of instructions (e.g., *look at the cat/cap box*) and asked to pick the word they had just heard between two options (e.g., *cat* or *cap*) on the screen while or after they listened to the input. Their eye movements over the visual screen while listening, along with their keyboard-press responses were recorded for the main analysis. The results showed both English and Korean listeners displayed higher proportions of fixations on the target (e.g., *cat*) than on the competitor words (e.g., *cap*) in assimilation contexts (e.g., *ca[^{t}p] box*), as well as higher proportions of fixations on targets (e.g., *cap*) than on competitors (e.g., *cat*) in non-assimilation contexts (e.g., *ca[p] box*). However, the discrepancy of fixation proportions between targets and competitors was more obvious for the English listeners than for the Korean listeners. In other words, although the L2 listenersin addition to L1 listeners were able to use acoustic variations when identifying the target phonemes, the L1 listeners revealed a higher certainty level than their L2 counterparts. Furthermore, the divergence points between targets and competitors wereshown to appear earlier for the L1 listeners than for the L2 listeners.

KEYWORDS

word recognition, assimilation, lexical ambiguity, L1 and L2 listeners, eye tracking, targets and competitors.

1. Introduction

Acoustic signals exhibit inherent variability that results from casual speech processes such as assimilation, reduction, and deletion. These processes lead to the drastic deviations of phonetic forms from the canonical pronunciation of the words intended by speakers. In English coronal place assimilation, coronal coda segments /t/, /d/, and /n/ are assimilated to the following bilabial or velar onset sounds, taking the place of the noncoronal segments. For example, *green berry* may sound like *greem berry*. This coronal place assimilation process is not obligatory, but this phenomenon is extremely common in natural English speech patterns (Gaskell and Marslen-Wilson 1996, Gow 2001, 2002, 2003, Darcy et al. 2007, 2009). The assimilation of the coronal place of articulation is also shown in other languages such as Korean. In Korean, the coronal coda consonants /t, n/ assimilate to the following bilabial sounds /m, p/ (e.g., /k’otʃ.pɑt/ [k’op.pɑt] ‘flower garden’, /ʃin.mun/ [ʃim.mun] ‘newspaper’) or the following velar sounds /k, ŋ/ (e.g., /sut.kɑrɑk/ [suk.k’ɑrɑk] ‘spoon’, /tʃən.kuk/ [tʃəŋ.kuk] ‘the whole country’). Although place assimilation may originate as a gradient phonetic effect triggered by acoustic characteristics or perceptual biases, both near-categorical and gradient dimensions of assimilation have been found in both English (Ohala1990, Byrd 1996, Gow 2001, among others) and Korean (Gow and Im 2004, Jun 1996, Kochetov and Pouplier 2008). Thus, place assimilation can be a source of lexical ambiguity as well as source of perceptual variability. For example, the phrase *ca[$^{t}_{p}$] box* can be perceived as either *cat box* or *cap box* when it is heard.

Spoken word recognition is dynamic and time-dependent. Listeners are generally good at identifying intended words despite phonetic modifications away from their canonical form. However, the non-canonical forms of words can sometimes cause difficulties in the recognition of words. The influence of fine-grained acoustic variability on L1 word interpretation has been repeatedly investigated (Beddor and Onsuwan 2003, Dahan et al. 2001, Gaskell and Marslen-Wilson 1996, Gow 2001, 2002, 2003, McMurray et al. 2002, 2008, Zamuner et al. 2016). On the other hand, previous studies demonstrating the effects of phonetic modifications on L2 listeners’ word recognition are still scarce. All languages manifest casual speech processes, and L2 listeners are not able to be permanently exposed to the input that are close to the canonical forms.

With this in mind, the current study investigated how L2 learners of English deal with English place assimilation that has the potential to cause lexical ambiguity. Specifically, this study explored the dynamics of lexical activation by comparing the time course of eye-movements of the two groups — native Korean and English listener groups — during real-time processes in assimilation and non-assimilation contexts (e.g., *ca[$^{t}_{p}$] box* vs. *ca[p] box*) in English. Additionally, we analyzed listeners' response data that include their judgments made based on their perception of the sound they heard. The specific research questions are as follows: 1. How does gradient modification in assimilation context influence word recognition? 2. How different are recognition patterns of phonetic information over time between L2-English listeners (native Korean listeners) and native English listeners?

In order to examine how gradient modification in an assimilation context influences word recognition, we used stimulus sets in which an assimilation process creates lexical ambiguity such as with *ca[$^{t}_{p}$] box* and *cap box*. Another issue is how phonetic information is integrated over time. To better understand the temporal dynamic of the perceptual processing of speech variation, we need experimental paradigms that provide continuous measures of activation. In this study, the Visual World Paradigm (VWP), an eye tracking method, was employed since it could reliably measure lexical activation with sensitivity to the temporal dynamics of processing.` In the following section, we will briefly review the relevant background on the assimilation and temporal processes of speech perception.

2. Literature review

2.1 Assimilation and Perception

As briefly addressed earlier, the assimilation process may produce lexical ambiguity since the process causes the neutralization of phonetic contrasts under certain environments (e.g., *ca[$^{t}{}_{p}$] box* vs. *ca[p] box*). In the perceptual aspect, listeners restore the phonological shape of words based on acoustic cues and phonetic contexts. Researchers have explored how listeners cope with phonological processes such as assimilation. Furthermore, a great deal of work has investigated context effects for assimilation focusing on the distinction between regular legal changes and abnormal deviations (Gaskell and Marslen-Wilson 1996, 1998, Gow 2001, 2002, Mitterer and Blomert 2003, Gow and Im 2004, Lee 2005, Darcy et al. 2009, Sung 2018). Gaskell and Marslen-Wilson (1996, 1998) examined context effects for place assimilation in English. Gaskell and Marslen-Wilson (1996) examined lexical access when the English coronal sounds /t, d, n/ were assimilated to either the bilabials /p, b, m/ or the velars /k, g, ŋ/ by using a lexical decision task. The English listeners demonstrated priming for the unmodified form of the target word (e.g., *wicked*) in the viable context in which assimilation is conditioned by a phonological process (e.g., *wicke[b] prank*) than in the unviable context in which assimilation is not triggered (e.g., *wicke[b] game*). Gaskell and Marslen-Wilson (1998) also found that surface variations in speech were perceptually tolerated when modifications occurred in phonologically viable contexts. The English listeners were able to access the mental representation of 'freight' easily when it was presented in the viable context (e.g., *freigh[p] bearer*) than in the unviable context (e.g., *freigh[k] bearer*). Furthermore, Gaskell and Marslen-Wilson observed the same tendency with nonce words, which suggests that phonological compensation occurred even without lexical cues.

Gow (2001) examined how listeners used the acoustic information in a token that had undergone assimilatory modification to anticipate the upcoming segment that triggered the assimilation. Gow found that when listeners heard the word *ten* in which the final coronal segment had been labialized, they expected that the next word would begin with a labial sound. For example, the word-initial noncoronal, as in the /b/ in *buns*, was perceived faster when it directly followed the properly modified underlying coronal, te[m] in *ten buns*, than when it directly followed the unmodified coronal, *te[n]* in *ten buns*. Gow pointed out that assimilation provides information about the underlying form as well as a particular type of modification.

The effects of language-specific experience concerning assimilation processes have been thouroughly explored. Lee (2005) investigated perceptual patterns involving Korean obstruent nasalization by Korean and English listeners. Lee showed that compared to the English listeners, the Korean listeners were highly sensitive to context and restored the phonological form for both Korean and English stimuli. That is, the Korean listeners revealed much higher word detection rates for the viable change condition (e.g., *pam.mul* 'rice water') than for the unviable change condition (e.g., *pap.panc^{h}an* 'a side dish'). Lee suggested that compensation for assimilation was clearly affected by L1 experience. Darcy et al. (2007) tested native English learners of French and native French learners of English with both their L1 and L2 through the involvement of assimilation. Each listener group consisted of two parties depending on their L2 proficiency. Darcy et al. found that beginners used their native compensation pattern in both languages, whereas advanced learners compensated for the non-native assimilation rule in their L2. Darcy et al. (2009) examined two phonological rules, one is the coronal place assimilation that exists in English and the other is the voicing assimilation that exists in French. The results of word detection tasks showed that both English and French listeners revealed a higher degree of compensation for phonological variations caused by rules existing in their L1 than by rules that were not present in their L1. Thus, the English listeners demonstrated more compensation for place assimilation, whereas the French listeners showed a higher compensation rate for voicing assimilation. It was also found that the non-native rule induced a small but significant compensation effect. Darcy et al. suggested that perceptual compensation involves both a language-specific and a language-independent mechanism.

Most of the previous studies on perception in an assimilation context have focused on perceptual compensation for phonological modification. The current study explored two issues related to the perception of the phonological form in an assimilation context. The first is how acoustic information involving assimilation affects word recognition. The second issue concerns the differences between L1 and L2 listeners in terms of the perception of acoustic variation over time. We tracked the activation of word recognition in a visual world task in which the participants matched spoken targets with the words on the computer screen.

2.2 Temporal Processes of Speech Perception

Eye-tracking has primarily been used to detect and measure one's eye movements (saccades) and stops (fixations), as well as the movement patterns over a text when reading (e.g., regressive or progressive patterns). Eye-tracking has also been utilized to explore the processing of auditory input. For example, a number of studies have investigated auditory word recognition and lexical activation using the Visual World Paradigm (VWP) (Gow and McMurray 2007, McMurray et al. 2008, Reinisch and Sjerps 2013, Schreiber and McMurray 2019, Desmeules-Trudel and Zamuner 2021, Reinisch and Mitterer 2022). In this procedure, the participants' eye movements are checked as they observe visual representations while listening to auditory stimuli. Eye-tracking methods have the advantage of providing the real-time processing of auditory information. By measuring the amount of time spent fixating on images or words on the display through a measurement of time, real-time lexical access as well as standard measures of language processing such as accuracy (e.g., mouse clicks) can be inferred.

Gow and McMurray (2007) examined the time-course of context effects using the VWP. The experiments employed items that produced potential lexical ambiguity caused by assimilation (e.g., the assimilation of /t/ in *cat box*). The results showed both progressive and regressive effects. Regarding the regressive effects, the participants favored images depicting *cats* when the context was labial (e.g., ca^t_p *box*) and *caps* when the context was coronal (e.g., ca^t_p *drawing*). That is, when the participants heard an assimilation token followed by a noncoronal (e.g., ca^t_p *,box*), their eye movements displayed that there was a small bias to prefer a noncoronal interpretation (e.g., *cap*) before 500 milliseconds (ms). However, at around 500ms, the preference reversed towards the coronal interpretation (e.g., *cat*). On the other hand, when they heard the same token in a coronal context (e.g., ca^t_p *drawing*), the fixation proportion of *cap* was higher than that of *cat* from around 500ms. Progressive effects were similar to those of regressive assimilation although progressive effects occurred earlier than regressive effects. Gow and McMurray suggested that perceptual inferences in the process of resolving assimilation might be bound up with the dynamics of lexical competition.

McMurray et al. (2008), Reinisch and Sjerps (2013), and Schreiber and McMurray (2019) investigated multiple cue integration across the time course, using the VWP. McMurray et al. (2008) provided evidence that suggests cues to both voicing (b/p) and manner (b/w) contrasts become available to listeners at different times during spoken word recognition. That is, the effect of the onset cue (VOT and formant transition slope) preceded the effect of vowel length. McMurray et al. pointed out that these results supported a model of cue-integration, and phonetic cues were used for lexical access as soon as they were available. Reinisch and Sjerps (2013) examined whether acoustic cues that are available on the same segment (duration and formant values in the Dutch contrast /a:/ vs. /ɑ/) would display differences in their use over time. Reinisch and Sjerps demonstrated that for a vowel contrast spectral cues tended to be used slightly earlier than duration cues although an equal importance of both cues was found in an offline task. Schreiber and McMurray (2019) tested anticipation by asking when listeners could use coarticulatory information in the frication to predict the upcoming vowel during speech perception. The sibilant fricatives /s/ and /ʃ/ were cross-spliced across vowels (e.g., the /s/ from *soup* and the /id/ from *seed*). Conditions in which the coarticulation matched or mismatched the vowel were compared (e.g., s_i*eed* vs. s_u*eed* and s_u*oup* vs. s_i*oup*). The VWP experiment found that listeners anticipated the vowel immediately from the onset of the frication, although they waited several hundred milliseconds to identify the fricatives. Schreiber and McMurray pointed out that listeners did not process phonemes in the order that they appeared, and suggested that dynamics of language processing may be loosely connected to the dynamics of the input.

Desmeules-Trudel and Zamuner (2021) investigated the effects of fine-grained phonetic details on L2 listeners' recognition of nasalized vowels in Canadian French. The results of two eye-tracking experiments indicated that L2 listeners (native English speakers) were able to use nasalization duration variability in a similar fashion to L1-French listeners. That is, L2 listeners were able to distinguish minimal word pairs differentiated by the presence of phonological vowel nasalization in French. Desmeules-Trudel and Zamuner argued that lexical representations could be highly specified in an L2. Furthermore, L2 listeners' knowledge of phonetic cues associated with phonological vowel nasalization in French depends on the age of exposure to the L2.

Previous studies using the VWP have yielded evidence of the real-time processing of auditory stimuli focusing on context effects in progressive and regressive assimilations, multiple cue integration, and L2 listeners' recognition of vowels. Thus, the VWP has provided a measure for tracking the use of phonetic cues over time. However, little research has been conducted regarding the effects of phonetic details on L2 listeners' real-time processing in an assimilation context. The present study compared the time course processing of phonetic details between L1 and L2 listeners. Specifically, this study compared perceptual patterns of phonological forms between an assimilation context (e.g., *cat* in $ca^t{}_p$ *box*) and a non-assimilation context (e.g., *cap* in *cap box*) throughout time. If complete assimilation is involved, the surface form of coronal sounds (e.g., $ca^t{}_p$) may be indistinguishable from the underlying noncoronal sounds (e.g., *cap*). However, it has been proven that assimilation is not a neutralizing process but rather a gradient incomplete modification (Gaskell and Marslen-Wilson 1996, 1998 Gow 2001, 2002 Gow and McMurray 2007). The appropriately assimilated form of coronal sounds (e.g., $ca^t{}_p$) may be perceptually distinguishable from the unmodified noncoronal sounds (e.g., *cap*). Accordingly, L2 listeners may notice fine-grained acoustic details in a way similar to that of L1 listeners.

3. Methods

3.1 Participants

A total of 34 participants, including twenty linguistically naïve native Korean speakers, as well as fourteen native English speakers, took part in this study. The Korean participants (4 males, 16 females) were recruited from large universities in Seoul, Korea. All of them were undergraduate students enrolled in different programs with their ages ranging between 20-29 years old (*M* = 24.45, *SD* = 1.77). Their English proficiency was measured via their self-rated proficiency using a 3-point Likert-scale (High-Intermediate-Low) and a voice recording of a short English paragraph. In the self-rated proficiency report, all except three participants (1 at low, and 2 at high) identified themselves as intermediate-level English learners. After confirming that the data from these three participants were not distinct from the rest of the group, we decided to include them in the analyses. Also, their English proficiency levels were considered to be intermediate or low based on the researchers' evaluation of their reading of the passage. The English participants (9 males, 5 females) were all residing in Seoul studying (*n* = 2) and working (*n* = 12) at the time of testing. Their ages varied ranging between 21 and 55 years old (*M* =37.36, *SD* = 9.92). Their length of residence in Korea also differed across participants, ranging between 3 months minimum to 12 years maximum (*M* = 7.24, *SD* = 3.69). Neither Korean nor English participants reported any hearing or vision problems. All the participants received monetary compensation for their participation.

3.2 Stimuli

Using the Experiment Builder (SR Research Ltd. ver.2.3.38) compatible with the Eyelink II eye-tracking system, the main listening task was designed by adopting the VWP so as to closely examine participants' time-sensitive proportional eye-gaze patterns between the two images as a function of their concurrent auditory input processing on the stimuli (for a more technical review of this method, see Altmann 2011, Tanenhaus and Trieswell 2006).

The stimuli submitted to the listening task consisted of a total of 190 test trials, including 10 practice, 120 experimental, and 60 filler items. The experimental items contained 30 items involving labialization (e.g., ca/t/ box → ca[p] box) and 30 counterparts with no labialization (e.g., ca/p/ box → ca[p] box), 30 items involving velarization (e.g., ba/t/ cage → ba[k] cage) and 30 counterparts with no velarization (ba/k/ cage → ba[k] cage). The experimental items were all compound nouns in the form of a target noun (one syllable) followed by a context noun (one or two syllables) as seen in Table 1. To avoid any potential confounding effect from semantic anomalies, all items were created so as to be as semantically plausible as possible. Then the auditory stimuli were recorded by a male native English speaker in the following sentence pattern: "Look at the EXPERIMENTAL ITEMS now." In this way, we were able to obtain more naturally occurring assimilated sounds in a natural utterance context. After the recording was over, the Praat software (version 6.1.03) was used to control the onset time of experimental item occurrences in that the first "Look at the" part was adjusted to 560 milliseconds across all items. Other than adjusting the onset time, all the auditory stimuli remained intact without any manipulations. The following table display examples of the stimulus items.

Table 1. Examples of Stimulus Items Involving Labialization and Velarization

Assimilation context		Target words		Context words
		Coronal sounds in assimilation contexts	Noncoronal sounds in non-assimilation context	
Labialization	[t] → [p] / __ [b], [p], [m]	cat	cap	box
	[d] → [b] / _ [b], [p], [m]	dad	dab	pose
	[n] → [m] / _ [b], [p], [m]	sun	some	bears
Velarization	[t] → [k] / _ [g], [k]	bat	back	cage
	[d] → [g] / _ [g], [k]	bid	big	game
	[n] → [ŋ] / _ [g], [k]	stun	stung	king

3.3 Procedure

All data collection processes took place in a quiet lab setting equipped with the Eyelink II eye-tracking system and a 23-inch LCD monitor screen for participant display. The experiment consisted of five pseudorandomized blocks with each block containing 35 to 40 trials mixed with the target and filler items and the participants were allowed to take a short break after completing each block.

Upon entering each trial, participants were instructed to fixate their gaze on the cross in the middle of the screen before receiving any visual stimulus, which was then followed by the audio prompt. They moved their eyes following the instructions in the prompt (i.e. by looking at the word that they believed they had just heard) and selected which word they heard by pressing the designated keyboard buttons. The number of times in which the target word appeared on the two sides (left and right) of the screen was counterbalanced to avoid any potential response biases.

While participants observed the images on the screen, they were asked to indicate the word they heard by pressing the left button if they heard the word on the left side, and the right button if they heard the word that appeared on the right side of the screen. In this manner, their proportional viewing of the two words during and after listening, together with their keyboard-click responses, were recorded for data analysis. The following figure displays a screenshot as an example of the experiment.

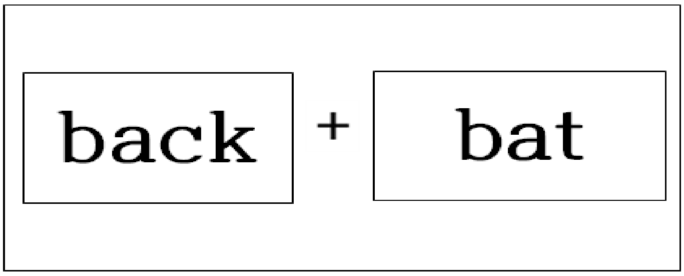

Figure 1. The Screenshot of the Experiment

4. Results

4.1 Keyboard-click Responses (Identification Task) 4.

The keyboard-click responses in the identification test were analyzed first. We assessed the rates for the selection of correct words with coronal codas (e.g., *cat*) in an assimilation context (e.g., *ca*[$^{t}_{p}$] *box*) and correct words with noncoronal codas (e.g., *cap*) in a non-assimilation context (e.g., *cap box*). The following figures show the response rates by English listeners (Figure 2) and Korean listeners (Figure 3).

(a) Assimilation context (b) Non-assimilation context

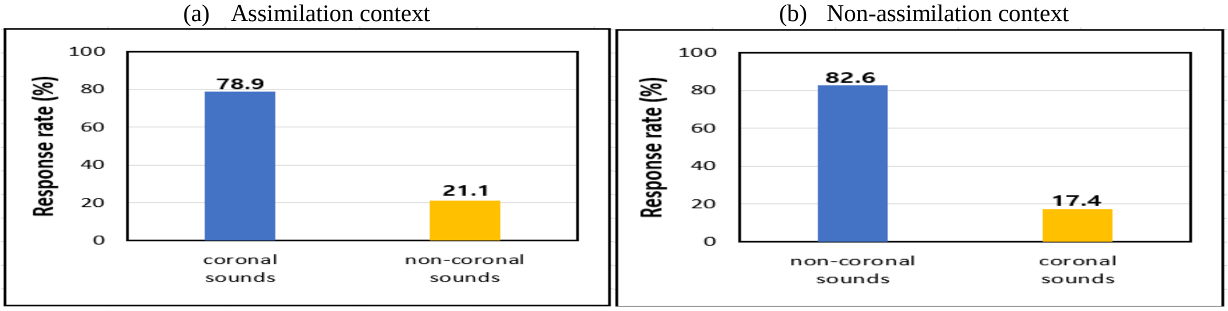

Figure 2. (a) English Listeners' Response Rates of Coronal Sounds (e.g., cat) and Noncoronal Sounds (e.g., *cap*) in an Assimilation Context (e.g., *ca[^{t}p] box*), and (b) Response Rates of Noncoroal (e.g., *cap*) and Coronal Sounds (e.g., *cat*) in a Non-assimilation Context (e.g., *cap box*)

The English listeners correctly identified around 78.9% responses of underlying coronal sounds in an assimilation context (e.g., *cat* in *ca[^{t}p] box*), and 82.6% of noncoronal responses in a non-assimilation context (e.g., *cap* in *cap box*). The noncoronal responses in the non-assimilation context were found to be less accurate than we had anticipated. This was likely due to the acoustic characteristics of coda consonants as coda consonants are not fully released in natural speech.

(a) Assimilation context (b) Non-assimilation context

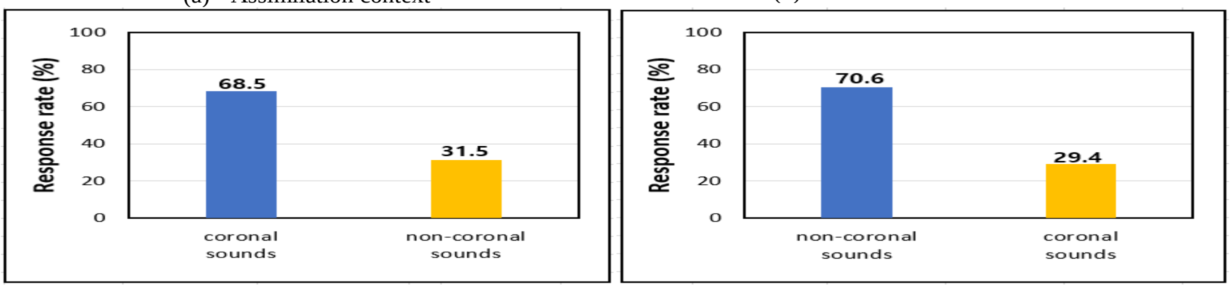

Figure 3. (a) Korean Listeners' Response Rates of Coronal Sounds (e.g., *cat*) and Noncoronal Sounds (e.g., *cap*) in an Assimilation Context (e.g., *ca[^{t}p] box*), and (b) Response Rates of Noncoronal Sounds (e.g., *cap*) and Coronal Sounds (e.g., *cat*) in a Non-assimilation Context (e.g., *cap box*)

The Korean listeners correctly identified around 68.5% of responses with coronal sounds in an assimilation context (e.g., *cat* in ca[^{t}p] box), and 70.6% of non-coronal responses in a non-assimilation context (e.g., *cap* in *cap box*). Compared with the results of the English listeners, the Korean listeners' detection rates of coronal sounds in an assimilation context and noncoronal sounds in a non-assimilation context were found to be lower. That is, the English listeners were more sensitive to the underlying form in both contexts.

A binary logistic regression was performed to ascertain the effects of the listener group (English vs. Korean listeners) and context (assimilation vs. non-assimilation context) on the responses (targets vs. competitors). The targets were coronal sounds in an assimilation context and noncoronal sounds in a non-assimilation context. The model explained 25% (Nagelkerke R^2) of the variances in responses and correctly classified 74.2% of cases. The results demonstrated that the listener group effects are significant (p = .000), and the effects of context are also significant (p = .044). The English listeners exhibited more target responses than the Korean listeners. Also, for both listener groups, the target responses were more clearly demonstrated in non-assimilation contexts than in assimilation contexts.

4.2 Eye Tracking Results

It has been well attested in literature that gaze fixations to visual targets reflect a viewer's attention which is essential for lexical activation in language processing (Tanenhaus et al. 2000). With this in mind, what interested us most in this study was 1) whether participants successfully detected the target sound both in the assimilated and non-assimilated environments, 2) if so, whether assimilated contexts in which the target sound was situated could somehow affect their lexical processing during listening, for example, with the timing of their sound recognition, and 3) whether there were discrepancies in the lexical processing between native English listeners and Korean listeners.

Figure 4 through Figure 7 present participants' time-course of looks (i.e. fixations) to the two words (e.g., *cat*, *cap*) in assimilation (e.g., *ca[^{t}p] box*) and non-assimilation (e.g., *cap box*) conditions, respectively. The time-course of the fixation in these figures demonstrates participants' proportional look on the two word images — namely the target and competitor words — from the time they heard the target word while listening to the audio prompt to the time they press the button that represented their response at the end.

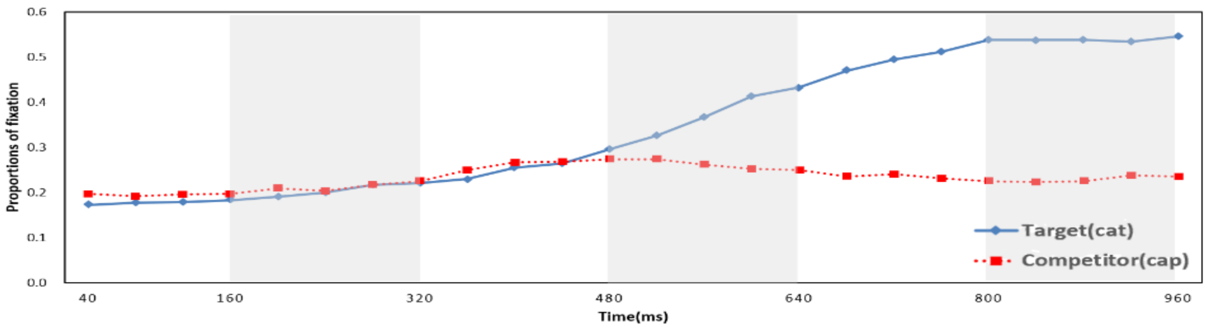

Figure 4. English Listeners' Proportions of Fixations on Coronal Targets (e.g., *cat*) and Noncoronal Competitors (e.g., *cap*) as a Function of Processing Assimilated Auditory Stimuli (e.g., *ca[^{t}p] box*)

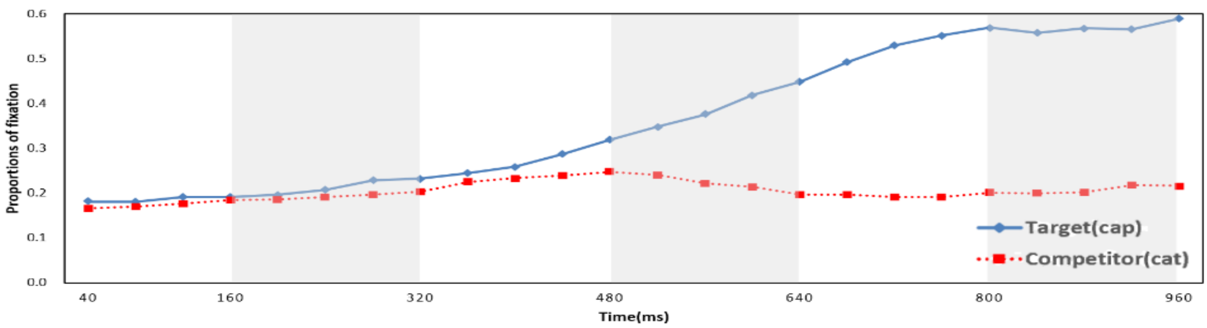

Figure 5. English Listeners' Proportions of Fixations on Noncoronal Targets (e.g., *cap*) and Coronal Competitors (e.g., *cat*) as a Function of Processing Non-assimilated Auditory Stimuli (e.g., *cap box*)

As shown in Figure 4 and Figure 5, the native English listener group showed similar fixation patterns in both conditions across the board. That is, they seemed to have clearly discriminated the target sound from the competitor in both conditions by gradually looking at the target word (i.e. the coronal target in assimilation, and the noncoronal target in non-assimilation conditions, respectively) more as time progressed. This observation conforms to the results of their keyboard click responses reported above in which their response accuracy ranged between 78.9 % in assimilation contexts and 82.6 % in non-assimilation contexts. However, the timing in which the proportions of the looks to the two fixation sites diverged slightly differed across the conditions. Specifically, English listeners' proportion of the looks to the target began to pile up and increase substantially from about 400 milliseconds after they heard the target words in non-assimilation conditions (see Figure 5), however, such a diverging point was observed at a later point in assimilation conditions, starting from about 480 milliseconds after they heard the target words (see Figure 4). This observed delay in assimilation conditions suggests that while English listeners processed the target sounds correctly in both conditions, processing assimilated coronals might have been more burdensome since they are phonologically somewhat ambiguous in nature. The following figures display the Korean listeners' fixation patterns.

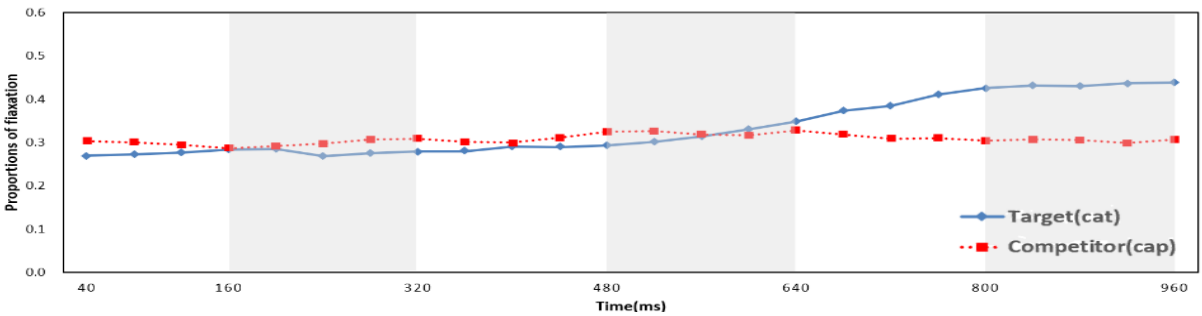

Figure 6. Korean Listeners' Proportions of Fixations on Coronal Targets (e.g., *cat*) and Noncoronal Competitors (e.g., *cap*) as a Function of Processing Assimilated Auditory Stimuli (e.g., *ca[^{t}p] box*)

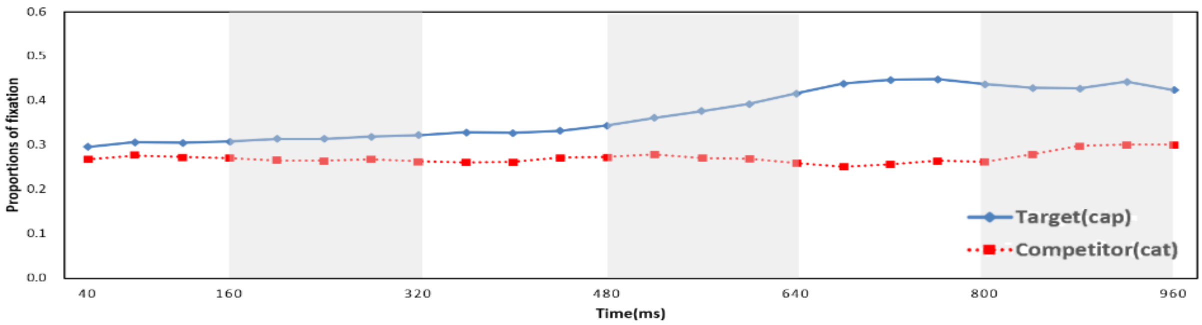

Figure 7. Korean Listeners' Proportions of Fixations on Noncoronal Targets (e.g., *cap*) and Coronal Competitors (e.g., *cat*) as a Function of Processing Non-assimilated Auditory Stimuli (e.g., *cap box*)

The Korean listeners also appeared to have processed the target words, by and large, correctly as evidenced by the increasing proportions of fixations on the target words in both conditions. See Figure 6 and Figure 7 for the fixation profiles in each of the two conditions. Comparing the two groups' time-course of fixations to the words, however, the Korean listeners displayed different profiles at least in two respects. First of all, although the Korean listeners eventually fixated their gaze more on the target words over their competitors as a function of time, the proportional differences between the target and competitor were much less, compared to the gaps seen in the English listeners' profiles. This could be partially due to their relatively lower response accuracy compared to the English listeners (68.5 and 70.6 percent in the assimilation and non-assimilation contexts, respectively). Furthermore, it also seems possible to speculate that the Korean listeners demonstrated relatively lower levels of certainty (or confidence) during the task even when they provided the correct responses toward the end of processing. Compared to the divergences between the targets and competitors shown by the English listeners (see Figures 4 and 5), the differences between the targets and competitors by the Korean listeners (see Figures 6 and 7) were much smaller. Second, the critical point at which the Korean listeners clearly discriminated the target sound from the competitor in the assimilation condition was found to have occurred at a much later point (about 640 milliseconds from the onset of the target words) in assimilation conditions, compared to the English listeners (about 480 milliseconds). This means that the English listeners were relatively faster than the Korean listeners in processing the assimilated target sound.

To further examine the time-course of fixations between the two groups, we conducted a series of statistical analyses. We first divided participants' time-course fixation data into six time-windows to compare time-course eye movement patterns of the two groups in these regions, each of which was set to include a 160-millisecond interval starting from the onset of the target words. We then calculated individual participants' proportional looks to the two sites — namely the targets and the corresponding competitors across experimental items — recorded during the span of each time-window. These proportional data were submitted to a series of generalized linear mixed-effects models (GLMMs) separately for each time window and for each condition, with Listener group (L1 English and L1 Korean) and Visual function (Target and Competitor) as the fixed factors, and with subject and item as random factors. Note that a significant interaction effect of the two fixed factors in the main analysis could signal that the two groups built different fixation patterns in some ways at the moment. To better identify the source of significance in interactions, if any, we performed sets of follow-up analyses separately each group using the GLMMs. The summary of the main analysis is provided in Table 2.

Table 2. The Effects of Listener Group (L1) and Visual Function

	Time region	1	2	3	4	5	6
	Elapse times in milliseconds	0-160	161-320	321-480	481-640	641-800	801-960
		F *P*	*F* *P*	*F* *P*	*F* *P*	*F* *P*	*F* *P*
L1	Assimilation	2.767 .096	1.699 .192	.380 .538	.001 .971	.405 .525	.369 .543
	Non-assimilation	3.139 .077	2.011 .156	.569 .451	.115 .735	.352 .553	1.437 .231
Visual Function	Assimilation	2.575 .109	1.644 .200	1.089 .297	26.061 **.000*****	168.859 **.000*****	243.165 **.000*****
	Non-assimilation	3.306 .069	9.729 **.002****	19.277 **.000*****	133.128 **.000*****	389.174 **.000*****	293.055 **.000*****
L1*Visual Function	Assimilation	.020 .888	.528 .468	.578 .447	22.605 **.000*****	45.629 **.000*****	38.599 **.000*****
	Non-assimilation	.699 .403	1.614 .204	.987 .321	5.918 **.015***	36.373 **.000*****	60.817 **.000*****

'***'<.001, '**'<.01, '*'<.05

First of all, the results displayed significant effects of visual function from Region 2 through Region 6 in non-assimilation conditions, largely due to substantially more looks to the target sites built from early stages for both groups. Similar results were obtained in assimilation conditions in this respect, however, a significant visual function effect appeared at later stages, starting from Region 4. Such delayed effects of visual function indicate that the participants did not likely conclude their processing of assimilated sounds up until Region 3. Second, the analysis found significant group by visual function interaction effects at Region 4 [Assimilation: $F(1, 4026) = 22.605$, $p < .001$, Non-assimilation: $F(1, 4030) = 5.918$, $p < .05$], Region 5 [Assimilation: $F(1, 3870) = 45.629$, $p < .001$; Non-assimilation: $F(1,3848) = 36.373$, $p < .001$], and Region 6 [Assimilation: $F(1, 3396) = 38.599$, $p < .001$; Non-assimilation: $F(1, 3338) = 60.817$, $p < .001$] in both conditions, respectively. In order to compare the processing of the two groups more closely during those timespans, follow-up analyses were carried out for each group, with the proportional looks as the dependent variables, and visual functions as the independent factor. A summary of the follow-up analyses is provided in Table 3.

Table 3. A Summary of Follow-up Analyses on Regions of 4, 5, and 6 for Each Group

	Assimilation Condition				Non-assimilation Condition			
Listener group	Korean		English		Korean		English	
	F	*P*	*F*	*P*	*F*	*P*	*F*	*P*
4	.073	.788	42.70	.000***	47.36	.000***	90.05	.000***
5	22.52	.000***	172.76	.000***	108.12	.000***	298.03	.000***
6	49.75	.000***	215.77	.000***	49.20	.000***	281.43	.000***

‘***’<.001, ‘**’<.01, ‘*’<.05

As shown in Table 3, one obvious difference between the English and Korean listeners came at time Region 4 in assimilations condition. That is, the English group had a significant effect related to the visual function during this time [$F(1, 1675) = 42.70$, $p < .001$] in which they were found to have looked at the target words considerably more. This suggests that they processed the given sound properly from the word strings during this moment of processing, consequently building up their fixations more on the target word site. In contrast, the Korean group did not display any effect related to the visual function during the same time span, [$F(1, 2352) = .073$, $p = .788$], and look probabilities on the target site did not much differ from those on the competitor site, reflecting some aspects of their uncertainty regarding the sound they had heard (see Figure 6). Taken together, the follow-up analyses revealed that the English listeners’ processing and recognition of assimilated sounds occurred at a relatively earlier stage of processing when compared to the Korean listeners. Other than the results at time Region 4 in the assimilation condition, the follow-up analyses did not find much statistically differing patterns despite the significant interactions of the two factors, in that both groups showed significant visual function effects across the regions (i.e. Region 4 through Region 6). This means that they discriminated the target sounds from the competitors properly during those time spans. Perhaps the significant interactions reported above might be due to different magnitudes of the gaps between the two groups’ looks to the target versus competitor words. Comparing their time-course fixation profiles (Figure 4 and Figure 6, and Figure 5 and Figure 7, respectively), it is apparent that the English listeners’ fixations on the target site were more stable than the Korean speakers across the board, which led their proportions of the look to the targets to diverge from the looks to the competitors in a more dramatic manner. For the Korean listeners, although they looked at the target site more, the proportional gaps between the targets and competitors were much less than those of the English listeners.

5. Discussion

The purpose of this study is to examine the dynamics of lexical activation when listeners hear the modified sounds in an assimilation context. For this purpose, we compared the time course of word recognition between coronal sounds in assimilation contexts and noncoronal sounds in non-assimilation contexts. This study also investigates different lexical activation patterns over time between native English listeners and Korean listeners learning English. We used eye tracking measures along with keyboard-click responses in order to examine processing stages.

The first research question of this study is how gradient modification in an assimilation context influences word recognition. The results of keyboard-click responses during the eye tracking experiment showed that both English and Korean listeners were able to recognize the underlying coronal sounds in an assimilation context. Further, both groups of listeners revealed higher response rates of underlying segments in a non-assimilation context than in an assimilation context. However, there was a discrepancy between the two listeners in the response rates. The English listeners' response rates of underlying coronal sounds in assimilation contexts were higher than those of the Korean listeners (78.9% vs. 68.5%). Also, the English listeners performed better than the Korean listeners in detecting the underlying sounds in non-assimilation contexts (82.6% vs. 70.6%).

While keyboard-click responses reveals the accuracy, the eye-tracking results provide a continuous measure of word activation over time. The present results of the eye-tracking task showed that both listener groups demonstrated different fixation proportions across the two conditions (assimilation vs. non-assimilation). The proportions of looks to targets began to substantially increase much earlier in the non-assimilation context than in the assimilation context for both English and Korean listeners (see Figures 4 – 7). Therefore, the influence of the assimilated form of coronal sounds on word recognition was clearly shown in both listener groups. Although the English listeners were more sensitive to acoustic cues than the L2 listeners, both listener groups were able to perceive fine phonetic details of coda sounds despite the incomplete release of codas. The present results are partially consistent with those of previous research (Mitterer et al. 2006, Desmeules-Trudel and Zamuner 2021). Mitterer et al. (2006) examined whether Dutch and Hungarian listeners compensated for Hungarian liquid assimilation. The authors found that compensation for assimilation can occur without experience with an assimilation rule in L1. Mitterer et al. argued that detecting underlying sounds depends on the phonetic details of assimilated segments, and that specific language experience is not required for compensation for assimilation at a prelexical level. Desmeules-Trudel and Zamuner (2021) examined how fine phonetic variability affected L2 listeners' recognition of nasalized vowels. Desmeules-Trudel and Zamuner pointed out that L2 listeners were able to utilize acoustic cues associated with vowel nasalization in a similar way to L1 listeners. In the present study we used the stimuli of English words involving coronal place assimilation. Although a clear divergence of word recognition in terms of timing and certainty level was observed between L1 and L2 listeners, acoustic variations triggered by assimilation were available for both listener groups.

The second research question involved asking how different recognition patterns of phonetic information are over time between L2-English listeners (Korean listeners) and native English listeners. We observed that both L1 and L2 listeners were able to use acoustic variability to detect underlying sounds. However, the results of the eye-tracking data also revealed different patterns of lexical activation between the two listener groups by measuring the proportions of looks to the targets and competitors. The Korean listeners' proportional discrepancy between the target and competitor was much less than that of the English listeners in both contexts as shown in Figures 4–7. Furthermore, in assimilation contexts, compared to the English listeners, the Korean listeners' discrimination points between the target and the competitor appeared much later. The English listeners' proportion of looks to the target diverged quite substantially from the competitor at around 480 ms, whereas the discrepancy was shown to occur around 640ms for the Korean listeners. These results demonstrated that although both L1 and L2 listeners could discern phonetic differences between assimilated coronal codas and noncoronal codas, L1 listeners recognized words with ambiguous codas earlier and with more certainty than L2 listeners.

These results are by and large consistent with those of previous studies that demonstrate that compensation for assimilation is affected by language-specific experience (Lee 2005, Darcy et al. 2007, 2009, Gow and Im 2004, Sung 2018). Lee (2005) found that compared to English listeners, Korean listeners were more sensitive to context involving obstruent nasalization, which exists in Korean phonology, but not in English. However, the English listeners also showed a marginal effect for words in the non-native process. Furthermore, Darcy et al. (2009) revealed that the English listeners were more sensitive to place assimilation that exists in English, whereas the French listeners were more sensitive to voicing assimilation that exists in French. Although the present study did not explore the context effects for compensation, the current results are generally in line with previous studies in terms of the effects of language-specific experience. The English listeners' detection rates of underlying forms in assimilation contexts were higher, and their recognition process was faster and contained more certainty than the Korean listeners. Thus, the present eye-tracking results demonstrated the discrepancies between L1 and L2 listeners regarding timing and robustness in the process of lexical activation.

6. Conclusion

By monitoring listeners' eye-movements during their real time processing of auditory input, this study examined how gradient modification in assimilation contexts influenced word recognition. We compared the time course of word recognition involving assimilated coronal codas and noncoronal codas in assimilation and non-assimilation contexts. We also compared fixation proportions of target and competitor words between L1 and L2 listeners. Overall, the results suggested some qualitative similarities between the two groups. That is, both listener groups demonstrated a higher proportion of fixations on targets than on competitors in both assimilation and non-assimilation contexts regarding phonetic details for the discrimination between assimilated coronal codas and noncoronal coda segments. However, the analysis of their time course data revealed some differences as well in terms of the timing of word recognition and the level of certainty. As discussed above, the English listeners' timing of word recognition was significantly faster than that of the Korean listeners. In addition, they demonstrated relatively greater and more stable focus on the targets compared to the Korean listeners. This could be taken to suggest that despite similar fixation profiles, the English listeners might have perceived the given sounds with more certainty and confidence during their real-time word recognition processes. We need more empirical research concerning the continuous variability of speech signals and word recognition. Eye-tracking methods can be used to examine the time course for inferring the upcoming segments in assimilation contexts. Furthermore, future work is needed to explore individual differences in terms of the timing and the certainty level in the process of lexical activation.

References

Altmann, G. T. M. 2011. The mediation of eye movements by spoken language. In S. P. Liversedge, I. D Gilchrist, and S. Everling, eds., *The Oxford Handbook of Eye Movements,* 974-1004. Oxford: Oxford University Press.

Beddor, P. S. and C. Onsuwan. 2003. Perception of prenasalized stops. In *Proceedings of the 15th International Congress of Phonetic Sciences,* 1, 407-410.

Byrd, D. 1996. Influences on articulatory timing in consonant sequences. *Journal of Phonetics* 24(2), 209-244.

Dahan, D., J. S. Magnuson and M. K. Tanenhaus. 2001. Time course of frequency effects in spoken-word recognition: Evidence from eye movements. *Cognitive Psychology* 42(4), 317-367.

Darcy, I., S. Peperkamp and E. Dupoux. 2007. Bilinguals play by the rules. Perceptual compensation for assimilation in late L2-learners. *Laboratory Phonology* 9(9), 411-442.

Darcy, I., F. Ramus., A. Christophe., K. Kinzler and E. Dupoux. 2009. Phonological knowledge in compensation for native and non-native assimilation. In F. Kügler, C. Féry & R. Vijver, ed.), *Variation and Gradience in Phonetics and Phonology*, 265-310. Berlin, New York: De Gruyter Mouton.

Desmeules-Trudel, F. and T. S. Zamuner. 2021. Spoken word recognition in a second language: The importance of phonetic details. *Second Language Research* Online First, https://doi.org/10.1177/02676583211030604

Gaskell, M. G. and W. D. Marslen-Wilson. 1996. Phonological variation and inference in lexical access. *Journal of Experimental Psychology: Human Perception and Performance* 22(1), 144-158.

Gaskell, M. G. and W. D. Marslen-Wilson. 1998. Mechanisms variation and inference in lexical access. *Journal of Experimental Psychology: Human Perception and Performance* 24(2), 380-396.

Gow, D. W. 2001. Assimilation and anticipation in continuous spoken word recognition. *Journal of Memory and Language* 45(1), 133-159.

Gow, D. W. 2002. Does English coronal place assimilation create lexical ambiguity. *Journal of Experimental Psychology: Human Perception and Performance* 28(1), 163-179.

Gow, D. W. 2003. Feature parsing: Feature cue mapping in spoken word recognition. *Perception & Psychophysics* 65(4), 575-590.

Gow, D. W. and A. Im. 2004. A cross-linguistic examination of assimilation context effects. *Journal of Memory and Language* 51, 279-296.

Gow, D. W. and B. McMurray. 2007. Word recognition and phonology: The case of English coronal place assimilation. *Papers in Laboratory Phonology* 9, 173-200.

Jun, J. 1996. Place assimilation is not the result of gestural overlap: Evidence from Korean and English. *Phonology* 13(3), 377-407.

Kochetov, A. and M. Pouplier. 2008. Phonetic variability and grammatical knowledge: An articulatory study of Korean place assimilation. *Phonology* 25(3), 399-431.

Lee, S. 2005. The effect of assimilation contexts in word detection. *Studies in Phonetics, Phonology and Morphology* 11(1), 105-124.

McMurray, B., M. K. Tanenhaus and R. N. Aslin. 2002. Gradient effects of within-category phonetic variation on lexical access. *Cognition* 86(2), B33-B42.

McMurray, B., M. A. Clayards., M. K. Tanenhaus and R. N. Aslin. 2008. Tracking the time course of phonetic cue integration during spoken word recognition. *Psychonomic Bulletin & Review* 15(6), 1064-1071.

Mitterer, H. and L. Blomert. 2003. Coping with phonological assimilation in speech perception: Evidence for early compensation. *Perception and Psychophysics* 65(6), 956-969.

Mitterer, H., V. Csépe and L. Blomert. 2006. The role of perceptual integration in the recognition of assimilated word forms. *Quarterly Journal of Experimental Psychology* 59(8), 1395-1424.

Ohala, J. 1990. The phonetics and phonology aspects of assimilation. In J. Kingston and M. E. Beckman, eds., Papers in *Laboratory Phonology I: Between the Grammar and Physics of Speech*. Cambridge: Cambridge University Press, 258-275.

Reinisch, E. and M. J. Sjerps. 2013. The uptake of spectral and temporal cues in vowel perception is rapidly influenced by context. *Journal of Phonetics* 41(2), 101-116.

Reinisch, E. and H. Mitterer. 2021. Phonetics and Eye-tracking. In R.A. Knight and Setter, eds., *The Cambridge Handbook of Phonetics*, 457-479. Cambridge: Cambridge University Press.

Schreiber, K. E. and B. McMurray. 2019. Listeners can anticipate future segments before they identify the current one. *Attention, Perception, & Psychophysics* 81(4), 1147-1166.

Sung, E. 2018. Compensation for phonological assimilation: Obstruent nasalization and coronal place assimilation. *Linguistic Research* 35(1), 145-178.

Tanenhaus, M. K., J. S. Magnuson., D. Dahan and C. Chambers. 2000. Eye movements and lexical access in spoken-language comprehension: Evaluating a linking hypothesis between fixations and linguistic processing. *Journal of Psycholinguistic Research* 29(6), 557-580.

Tanenhaus, M. K. and J. C. Trueswell. 2006. Eye movement and spoken language comprehension. In M. J. Traxler and M. A. Gernsbacher, eds., *Handbook of* Psycholinguistics, 835-862. Amsterdam: Elsevier.

Zamuner, T. S., E. Morin-Lessard., S. Strahm and M. P. Page. 2016. Spoken word recognition of novel words, either produced or only heard during learning. *Journal of Memory and Language* 89, 55-67.

Appendix

1. Stimulus examples (labialization)

Assimilation context	Target words		Context words
	Coronal sounds in assimilation contexts	Noncoronal sounds in non-assimilation contexts	
[t] → [p] / __ [b], [p], [m]	mat	map	paper
	cat	cap	box
	beat	beep	maker
[d] → [b] / _ [b], [p], [m]	dad	dab	pose
	grad	grab	ball
	dud	dub	machine
[n] → [m] / _ [b], [p], [m]	warn	warm	people
	sun	some	bears
	teen	team	member

2. Stimulus examples (velarization)

Assimilation context	Target words		Context words
	Coronal sounds in assimilation contexts	Noncoronal sounds in non-assimilation contexts	
[t] → [k] / _ [g], [k]	bat	back	cage
	bait	bake	guide
	shot	shock	gun
[d] → [g] / _ [g], [k]	bud	bug	killer
	bid	big	game
	rid	rig	gas
[n] → [ŋ] / _ [g], [k]	thin	thing	counter
	stun	stung	king
	ban	bang	games

Examples in: English
Applicable Languages: English
Applicable Level: Tertiary