Journal of Memory and Language 69 (2013) 59–83

Contents lists available at ScienceDirect

Journal of Memory and Language

journal homepage: www.elsevier.com/locate/jml

Compensation for complete assimilation in speech perception: The case of Korean labial-to-velar assimilation

Holger Mitterer [a], Sahyang Kim [b], Taehong Cho [c,*]

[a] Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands
[b] Department of English Education, Hongik University, Seoul, Republic of Korea
[c] Hanyang Phonetics and Psycholinguistics Lab, Department of English Language and Literature, Hanyang University, Seoul, Republic of Korea

ARTICLE INFO

Article history:
Received 17 February 2011
Received in revised form 8 August 2012
Available online 26 March 2013

Keywords:
Spoken word recognition
Place assimilation
Compensation
Korean
Dutch
English

ABSTRACT

In connected speech, phonological assimilation to neighboring words can lead to pronunciation variants (e.g., 'garden bench' → "garde*m* bench"). A large body of literature suggests that listeners use the phonetic context to reconstruct the intended word for assimilation types that often lead to incomplete assimilations (e.g., a pronunciation of "garden" that carries cues for both a labial [m] and an alveolar [n]). In the current paper, we show that a similar context effect is observed for an assimilation that is often complete, Korean labial-to-velar place assimilation. In contrast to the context effects for partial assimilations, however, the context effects seem to rely completely on listeners' experience with the assimilation pattern in their native language.

Introduction

The acoustic–phonetic form of a given word is by nature variable as it is conditioned not only by inevitable consequences of physiological differences between speakers but also by other factors that are under speakers' control, such as speaking rates and prosodic manipulation. Another important factor that influences shaping the acoustic–phonetic form of a word is its phonetic context—i.e., what segments come before and after the word, as they often trigger coarticulation or assimilation between segments across the word boundary (see Kühnert & Nolan, 1999). This influence can be so strong that phonemic distinctions between sounds may be blurred under the contextual influence. For instance, when the German word "Wälder" (/vɛldɐ/, 'forests') that starts with a voiced consonant /v/ is produced after a word that ends with a voiceless obstruent (e.g., /t/) as in "... hat Wälder..." ('...has forests'), it may sound similar to the word starting with /f/, "Felder" (/fɛldɐ/, 'fields'), due to a progressive voicing assimilation—i.e., the voicelessness of the preceding /t/ leads to devoicing of the following /v/ (Kuzla, Ernestus, & Mitterer, 2010). The assimilatory change is, however, often more subtle than this abstract description portrays. It often leads to gradient phonetic segments that are "in between" the underlying segment (e.g., /v/) and the categorically assimilated segment (e.g., [f]). Kuzla et al. (2010) showed that the word-initial /v/ can become fully devoiced after /t/ but its fine-grained phonetic detail is still different from that of an underlying /f/. The assimilated /v/ maintains the durational characteristics of its underlying /v/—that is, it is shorter than the underlying /f/.

There is a large body of studies that examines how listeners cope with such changes, especially focusing on the alteration of place of articulation of word-final alveolar segments (e.g., in "garden bench" /gɑ:dn bɛntʃ/ → [gɑ:

* Corresponding author. Address: Hanyang Phonetics and Psycholinguistics Lab, Department of English Language and Literature, Hanyang University, 17 Haengdang-dong, Seongdong-gu, Seoul 133-791, Republic of Korea. Fax: +82 2 2220 0741.
E-mail addresses: holger.mitterer@mpi.nl (H. Mitterer), sahyang@hongik.ac.kr (S. Kim), tcho@hanyang.ac.kr (T. Cho).

0749-596X/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.jml.2013.02.001

dmbɛntʃ]; see Coenen, Zwitserlood, & Bölte, 2001; Darcy, Peperkamp, & Dupoux, 2007; Gaskell & Marslen-Wilson, 1996; Gaskell & Snoeren, 2008; Gow, 2003; Mitterer & Blomert, 2003). These studies have converged on the conclusion that listeners make use of the contextual information to compensate for such assimilatory changes. That is, a changed word form is recognized "correctly" only in the context that licenses the assimilatory change. For example, an assimilated word form "garde[m]" is recognized as intended ('garden') only if it is followed by a word that starts with a bilabial sound (e.g., /b/ in 'bench') that triggers the assimilation. Such a viability effect has been observed for English coronal place assimilation (Gaskell & Marslen-Wilson, 1996, 2001; Gaskell & Snoeren, 2008; Gow, 2003), nasal place assimilation in Dutch and German (Coenen et al., 2001; Mitterer & Blomert, 2003; Tavabi, Elling, Dobel, Pantev, & Zwitserlood, 2009), voicing assimilation in French (Darcy et al., 2007; Snoeren, Segui, & Hallé, 2008), and manner assimilation in Hungarian (Mitterer, Csépe, & Blomert, 2006; Mitterer, Csépe, Honbolygo, & Blomert, 2006).

There is, however, considerable disagreement over the processing mechanism that drives this context-driven viability effect. For example, Gaskell (2003) proposed a statistical learning model that assumes that listeners learn which phonological alterations occur in which contexts in a given language. Through this learning, listeners build up statistically-informed phonological knowledge. While the model was initially developed based on the assumption that phonological changes are categorical (Gaskell & Marslen-Wilson, 1998), it has been changed to allow for gradient phonetic manifestation of phonological feature changes due to assimilation (Gaskell, 2003). Gow's (e.g., 2003) feature-parsing account, on the other hand, gives more weight to the role of language-independent perceptual-grouping processes in accounting for the recognition of assimilated words. It crucially assumes that assimilation is phonetically partial and gradient, leaving the information of the underlying representation of the assimilated segment, so that listeners exploit residual phonetic detail not only in recovering its underlying representation but also in predicting the upcoming segment that triggers the partial assimilation.

Mitterer and colleagues (Mitterer, Csépe, & Blomert, 2006; Mitterer, Csépe, Honbolygo, et al., 2006) proposed yet another account, which they called the perceptual-integration account. According to this account, both language-independent and language-dependent factors play a role. The language-independent part of the account is inspired by functional models of phonetically-driven phonology (Blevins, 2004; Boersma, 1998; Steriade, 2001), which assume that basic auditory abilities have a role to play in shaping phonological alterations in production. The fundamental assumption is that some sounds are perceptually more salient than others, such that the perceptually more robust sounds tend not to undergo phonological alterations for the listeners' benefit whereas perceptually weaker sounds tend to be phonologically modified, driven by the principle of ease of articulation. (See also Cho and McQueen (2008) and Hura, Lindblom, and Diehl (1992) for evidence for differential perceptual stability among different manners of articulations in line with this assumption.) Mitterer and colleagues extended this concept to include context effects: If a segment is difficult to be perceived in a given context, that segment is likely to be assimilated. This assertion was based on their findings on Hungarian manner assimilation. They showed that a change from /l/ to /r/ in Hungarian is difficult to perceive in front of another /r/ for both Hungarian and Dutch native listeners. This perceptual instability, driven by the language-independent interaction of the perceptual robustness and the ease of articulation, was arguably why /lr/ sequences across a word boundary are produced as [rr] in Hungarian (i.e., assimilation of /l/ to /r/).

While the perceptual-integration account focuses on this interplay of perception and production, giving rise to language-independent context effects, it also includes an auxiliary assumption to explain effects of language-specific phonological knowledge. For example, Mitterer, Csépe, and Blomert (2006) and Mitterer, Csépe, Honbolygo, et al. (2006) found that Hungarian listeners were more likely to categorize the ambiguous stimuli [Xr] (where 'X' refers to a perceptually ambiguous sound between /l/ and /r/) as /lr/ than Dutch listeners were, showing their use of language-specific phonological knowledge (i.e., assimilation of /l/ before /r/) in recovering the underlying phoneme of an ambiguous sound. That is, when hearing [Xr] sequence, Hungarian listeners, knowing that /lr/ is produced as [rr] in the language, may have learned to interpret it more likely as [lr], while non-native (Dutch) listeners have no such language experience.

A similar mixture of language-dependent and -independent effects was also observed in the fricative assimilation in English (e.g., in 'glass shop' /glɑs ʃɒp/ → [glɑʃːɒp]) by English listeners who are familiar with the process versus French listeners who are not (Gaskell, Clayards, & Niebuhr, 2009). (See also Darcy et al. (2007) for relevant findings.)

What appears to emerge from some of the existing data is that the phonological identity of the assimilated segment may be blurred as a language-independent phonetic consequence of assimilation processes, but language-specific phonological knowledge helps in resolving the ensuing ambiguity correctly. However, an important issue is the extent to which the phonetic detail plays a role in compensation for assimilation. As discussed above, the feature parsing account (Gow, 2002, 2003) depends on the phonetic detail coming from incomplete or partial assimilation, which does not require language-specific phonological knowledge. It also predicts that phonological knowledge is not invoked even when assimilation is complete, because both native and non-native listeners would have no residual acoustic–phonetic cues to use in recovering the underlying representation.

Gow and Im (2004) indeed assessed the role of completeness of assimilation and language experience at the same time, especially by investigating the perception of Hungarian regressive voicing assimilation and Korean regressive labial-to-velar place assimilation. Critically, the measurements of their stimuli showed that the Hungarian regressive voicing assimilation in their stimulus set were partial while the Korean labial-to-velar place assimilation in their stimulus set were complete. Based on the feature-parsing account, they tested whether participants would be faster to detect a segment if it triggered assimilation.

The logic was as follows. If participants are able to predict the upcoming segment which triggers assimilation based on the acoustic–phonetic information of the assimilated segment that is currently being processed, the listeners must have used a feature parsing mechanism, by associating the phonetic information of the partially assimilated segment with the assimilation-triggering segment that follows. The underlying representation of the assimilated segment is also recovered as intended because the phonetic difference between the underlying and assimilated segments is attributed to the post-assimilation context. In line with this assumption, the detection of the post-assimilation segment (that triggered assimilation) was found to be facilitated in the incomplete voicing assimilation environment in Hungarian, but not for Korean labial-to-velar assimilation, which was assumed to be a complete process. (At least in their stimuli, no phonetic residuals existed.)

To test whether these results were modulated by language experience, English listeners performed the same task as the native Hungarian and Korean listeners. Crucially, it turned out that both findings were independent of language experience—i.e., both Hungarian and English native listeners, with Hungarian speech stimuli, showed comparable facilitation effects while both Korean and English native listeners, with Korean speech stimuli, failed to predict the upcoming velar based on the derived velar. Based on these results, they argue that language-specific phonological knowledge plays little role in coping with assimilation-driven variability, regardless of whether there is a bottom-up acoustic–phonetic support for the underlying representation (as in the partial assimilation case in Hungarian) or not (as in the complete assimilation case in Korean labial-to-velar assimilation).

While Gow and Im (2004) suggested that there is no context effect in complete assimilation, some recent studies have challenged this theory. Gaskell and Snoeren (2008), for example, were able to obtain natural examples of assimilations (*run picks* produced as "rum picks") that were at least perceptually complete in a production study. In contrast with what the feature parsing account predicts, they found that these complete assimilations were still subject to context effects: Listeners were more likely to interpret a completely assimilated word as ending in an alveolar sound in an assimilation-triggering context than in the control context.

More directly related to Gow and Im's results on labial-to-velar assimilation in Korean is a phoneme monitoring study by Cho and McQueen (2008) which showed differential perceptual consequences of alveolar-to-velar versus labial-to-velar place assimilation in Korean. In line with previous findings on the alveolar assimilation, they found that alveolar-to-velar assimilation hardly burdens the listeners. In their experiment, participants were equally fast in detecting an alveolar segment regardless of whether the specified target phoneme was physically present (unassimilated) or absent (assimilated). However, when it comes to a labial target, they found it harder to detect the specified target (i.e., labial) when it was not physically present (assimilated). Most crucially, in creating the speech materials, Cho and McQueen used completely assimilated forms, so that there were no acoustic–phonetic residual cues for the underlying targets in both assimilation cases. Nevertheless, listeners were able to perceive intended segments successfully in both cases, but the processing burden was not the same in recovering alveolar targets and labial targets, suggesting a possibility that different types of assimilation are processed by different mechanisms. This contrasts with one of the underlying assumptions of the feature parsing theory which predicts that any kind of regressive assimilation is processed similarly through language-independent auditory-perceptual mechanisms.

Mitterer (2011b) also presented a case in which compensation for very similar reduction processes may be achieved via different mechanisms. In Dutch, /b/ can trigger either place assimilation as in /tœynbɑŋk/ → [tœymbɑŋk] ('garden bench') or reduction of a word-final /t/ as in /mɛst bəstɛlt/ → [mɛsbəstɛlt] ('fertilizer ordered'). Parsimony would seem to dictate that both context effects arise as the consequence of the same processing mechanism. The results of a discrimination task, however, indicated that only the results of nasal place assimilation, but not the results of /t/-reduction, were partly compensated for by early auditory processes.

Given that the results in Cho and McQueen (2008) imply that "not all sounds in assimilation environment are perceived equally" and given that compensation for different forms of reduction may differ (Mitterer, 2011b), understanding how listeners cope with labial-to-velar assimilation in Korean is a crucial piece of the puzzle. Most importantly, there are not enough studies in the literature that indicate how listeners deal with a categorical (or complete), but optional assimilation rule like the Korean labial-to-velar assimilation case—i.e., a type of assimilation that may not always occur, but is always complete if it occurs. In Korean, unlike in English or Dutch, both alveolars and labials can be optionally assimilated in place to the following velar (e.g., Kim-Renaud, 1975; Jun, 1996). This optional assimilation is not triggered simply by the phonological context, but it is likely to occur in casual speech, especially when the target and the trigger consonants occur without a major phrase boundary that intervenes the consonant sequence. Recent phonetic studies (e.g., Gow & Im, 2004; Son, Kochetov, & Pouplier, 2007) have suggested that assimilation is categorically complete if it occurs, showing no gradient residual phonetic cues to underlying forms. It is therefore worth pursuing whether the mechanisms for compensation for a categorical type of assimilation differ from the mechanisms for compensation for a gradient type of assimilation.

The purpose of the present study is, therefore, to test whether the context plays a role in compensation for Korean labial-to-velar assimilation, which has been known to be optional but complete. Consider the situation in which a listener hears a word ending on a velar (e.g., [tʃʰaŋ], 'window') even though the intended word actually ended on a labial (i.e., /tʃʰam/, 'truth'). Are listeners more likely to overcome this mismatch if it occurs in an assimilation-triggering context (e.g., [tʃʰaŋ#k...]) than in a control context (e.g., [tʃʰaŋ#s...])? The phonological-inference account (Gaskell, 2003) predicts such a context effect, because

Korean listeners should have acquired language-specific phonological knowledge about the assimilation in their native language. The perceptual-integration account (Mitterer, Csépe, & Blomert, 2006) also predicts a context effect, but this context effect should arise on an auditory level of processing and hence be independent of phonological knowledge. Note that the perceptual-integration account does not require assimilations to be incomplete; it only assumes that the assimilation has minimal perceptual consequences (cf. Hura et al., 1992). so that the assimilated version sounds very similar to the unassimilated version with the latter being perceptually integrated with the percept of the following contextual sound. In this regard, it differs from the feature-parsing account, for which the presence of residual cues for the underlying segment is crucial. Therefore, the feature parsing account (Gow, 2003), in contrast, predicts no context effect for tokens of complete assimilation.

The present study also has an important methodological motivation by conducting a spoken word recognition task in an eye-tracking paradigm, following Mitterer and McQueen (2009) that demonstrated the effectiveness of the eye-track paradigm in testing context effects. Mitterer and Ernestus (2006) showed that, in Dutch, /t/ reduction is more likely to occur if the following word starts with /b/ than when it starts with /n/. They tested the context effect by asking listeners to indicate whether a nonword ended with /t/ or not, in /b/ versus /n/ contexts in a two-alternative forced choice (2AFC) task. But they failed to find an effect of the following context in /t/-reduction. Subsequently, Mitterer and McQueen (2009) tested again whether this null effect of the following context was indeed due to the true absence of a context effect or due to the meta-linguistic nature of the 2AFC task. This time, they employed the visual world eye-tracking paradigm, in which participants heard a sentence and saw several printed words (usually four) on the screen (McQueen & Viebahn, 2007). The displays contained minimal pairs differing in the presence of a word-final /t/ (e.g., /kʏs/–/kʏst/, 'kiss'–'coast') above or next to different geometrical shapes (a circle, a star, a triangle, or a rectangle). With these displays, participants heard instructions to click on one of the printed words above or next to a certain shape (e.g., "click on the word *kiss* above the star").

There are four things to note about this type of instruction. First of all, the potential overspecification of the target by adding the shape is not unnatural. Especially in situations in which instructions are given with no direct feedback (as in this experimental situation), speakers often make use of overspecification (Maes, Arts, & Noordman, 2004). Second, the instruction is ambiguous because the Dutch words for "kiss" and "coast" can both be produced as [kʏs] in connected speech. Third, this ambiguity is resolved at the last word of the instruction which indicates with which shape the target word was paired (in this example, only one of the two candidate words was positioned above a star). Hence, the participants were not forced to make a meta-linguistic decision. Fourth, and most importantly, the printed words on the screen could be above or next to the geometrical shapes and the Dutch word for 'above' starts with /b/ (/bovə/) while the Dutch word for 'next to' starts with /n/ (/nast/), allowing for testing /b/ versus /n/ context effects. According to the production data in Mitterer and Ernestus (2006), the reduction of /t/ is more likely to occur before /b/ in the sentence with "above" than before /n/ in the sentence with "next to" following the target word. Crucially, Mitterer and McQueen found that listeners indeed took this into account: There were more looks toward the word with /t/ if it was followed by /bovə/ than when followed by /nast/.

The bottom-line is that context effects, which were not observed in an otherwise sensitive categorization task (Mitterer & Ernestus, 2006), can be observed in the eye-tracking paradigm with printed words and geometrical shapes. In the present study, we employ this paradigm to revisit context effects in perception of Korean labial-to-velar assimilation by focusing on how listeners perceive the completely assimilated segments, different from Gow and Im (2004) who failed to find the effect of completely assimilated segments on perception of the post-assimilation segments.

Before moving into the next section for Experiment 1, it is worth addressing some concerns with respect to the use of printed words in the experiment. In an eye-tracking experiment with English stimuli, Salverda and Tanenhaus (2010) showed that competitors (e.g., *bear* vs. *bare*) with some degree of orthographic overlap with the target (e.g., *bead*) received more looks than unrelated distractors (e.g., *gild*, *frog*). Crucially, when the critical competitors were homophonous (e.g., *bear* vs. *bare*) and hence had the same degree of phonological overlap with the target (e.g., *bead*), the competitor with a higher degree of orthographic overlap with the target (e.g., *bear* with *bead*) received more looks than the one with a less orthographic overlap (e.g., *bare* with *bead*). These results, as the authors concluded, suggest that the mapping between spoken words and printed words is mediated by orthographic representations upon hearing spoken words. It is therefore important to assure that the present study (and any study using the printed-words version of the visual-world paradigm) is not confounded with such orthographic similarity effects.

The present study with Korean is indeed most likely to be free from such confounds for the following reasons. First, unlike the linear alphabetic systems of Indo-European languages (including English), the Korean orthographic system, *Hangul,* is a non-linear alphabetic system. In *Hangul*, letters in a syllable are arranged in a syllable block in a combinatory left-to-right and top-to-bottom fashion as shown in Fig. 1 (see Lee & Ramsey, 2000, p. 13, for a complete description of the Korean writing system). It is therefore unlikely that the kind of orthographic similarity effects observed in English applies to Korean.[1] Second, *Hangul* was invented (rather than developed) as a phonologically transparent writing system (with a clear letter-to-sound mapping), so that the underlyingly heterographic homophones do not exist in

[1] It is in fact an open question whether the orthographic-similarity effect reported by Salverda and Tannenhaus is generalizable to writing systems with more transparent phonology-to-orthography relationships than the one in English. The exceptional status of English orthography is evident in reading development, in which English children need 4 years of teaching to reach the same reading level achieved after 1 year by children faced with more transparent orthographies, see Aro and Wimmer (2003).

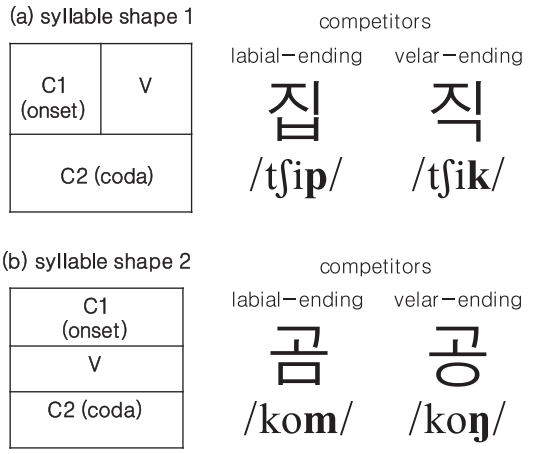

Fig. 1. Examples of printed words in Korean using the Hangul script. Panels (a) and (b) show two different possibilities for the spatial arrangements in syllable blocks in Hangul.

Fig. 2. Experimental display as used in Experiment 1, with phonetic transcriptions (in gray) added for the readers' convenience.

Korean (although apparent phonological neutralization may take place due to phonological rule applications such as assimilation). The orthographic representations are, therefore, by and large predictable from pronunciations. Finally, and most importantly, as shown in Fig. 1, the orthographic visual forms of competitors that are to be used in the present study always share the identical onset and vowel, and they differ only for the critical final letter (either a labial or a velar consonant; see the next section for more details) positioned in the bottom coda slot (C2 in Fig. 1). Given the phonological transparency of the Korean writing system, and given the same degree of orthographic overlap between competitors used in the present study, the possibility of the orthographic confounding effects on the task should be minimal.

Experiment 1

For the current experiment, we adapted the paradigm used in Mitterer and McQueen (2009), as shown in Table 1 and Fig. 2. Table 1 shows the frame of auditory instruction given to the participants in the experiment. The critical word is from a minimal pair with a labial or velar final consonant (e.g., '집' /tʃip/ – '직' /tʃik/), which is followed by either the control context with an /s/-initial word (i.e., /saŋtanɛ/, 'above'), or the assimilation-triggering context with a /k/-initial word (i.e., /kak*ai/, 'near'). Fig. 2 is an example of screen display with four printed words in Hangul character strings and four geometrical shapes (a rectangle, a star, a triangle, and a circle).

Following the examples of Mitterer and McQueen (2009) and Mitterer (2011a), we had two different types of trials for the experiment: phonological trials and semantic trials. In phonological trials, participants see four different printed words on the screen which constitute two phonological minimal pairs, with contrasts shown in word-final segments (i.e., labial versus velar as in 집 /tʃi**p**/ – 직 /tʃi**k**/; 참 /tʃʰa**m**/ – 창 /tʃʰaŋ/), as in Fig. 2. Recall that the two minimal pairs (집 /tʃi**p**/ – 직 /tʃi**k**/ and 참 /tʃʰa**m**/ – 창 /tʃʰaŋ/) differ both orthographically and phonologically only in the final consonant which is located at the bottom of the syllable block. In these phonological trials, spoken words ending with a velar are critical. When a velar-ending test word appears in the assimilation-triggering velar context (i.e., before /k/-initial word as in [tʃi**k #** kak*ai] where '#' is a word boundary), it can be considered either as an original

Table 1
Sentence frame for the instructions, translated into a English as "Click on the [star|circle|rectangle|triangle] above/next to which is *TARGET* on the screen".

	Preceding context	Target	Following context 1	Geometrical shape	Following context 2
Korean	화면에서			별	
			상단에 있는	원	을(를)
			가까이에 있는	네모	누르세요
				세모	
Phonetic form	hwamjʌn-ɛsʌ	(CVC).(C)V**k**	saŋtan-ɛ-itn̩in	pjʌl	ɨl (lɨl)-
		(CVC).(C)V**p**		wʌn	
		(CVC).(C)Vŋ	kak*ai-ɛ-itn̩in	nɛmo	nurɨ-sɛjo
		(CVC).(C)V**m**		sɛmo	
Gloss	Screen-from		Above-LOC-exist	Star	ACC-
				Circle	
			Near-LOC-exist	Rectangle	Click
				Triangle	

velar-ending word (with underlying /k/ or /ŋ/), or as an assimilated velar-ending word (i.e., derived from underlying labial /p/ or /m/). The phonological trials are therefore designed to test the phonological context effects (i.e., whether the assimilation-triggering context would influence listeners' perception of preceding words that end with either a velar or a labial), which is the main goal of Experiment 1. Critically, this experiment was designed in such a way that listeners were often asked to accept a labial-ending *printed* word (e.g., '**집**'/ tʃi**p**/) upon hearing a velar-ending *spoken* stimulus [tʃi**k**]. This situation arose when the matching written word was accompanied by a different shape than the one used in the instruction sentence. We expected that, in such cases, participants would click on the shape mentioned in the instruction sentence which was 'above' or 'next to' the mismatching word. If listeners take the phonological context into account, they should find it easier to accept the mismatch (i.e., the labial-ending printed word versus the velar-ending spoken word) in the assimilation-triggering (/k/-initial) context than in the control (/s/-initial) context.

In semantic trials, it was measured how the semantic information in the auditory instruction sentence guided the listeners' task. The critical semantic information was either the relative position of the object to be clicked on (e.g., 'above' or 'next to' the printed word) or the name of the geometrical shape to be clicked on (e.g., a star or a circle). In semantic trials, participants also see four printed words on the screen, but in these trials, the four printed words are two pairs of identical words. That is, two printed words appear twice on the screen in semantic trials, unlike the display of four distinctive printed words in phonological trials as depicted in Fig. 2. These trials are nevertheless unambiguous because two pairs of identical words are disambiguated by semantic information in the auditory instruction sentence that listeners hear for each trial. Depending on the source of disambiguating information, semantic trials are further divided into *Position* trials and *Shape* trials. In *Position trials* (with the auditory instruction sentence with the specified position of the object relative to the printed word), two identical *Hangul* character strings (i.e., two identical printed words) are accompanied by the same shape, but the shape is in different positional relation with the test words on the screen (i.e., positioned either "above" the word or "near" (next to) the word). Disambiguating information is then auditorily supplied in the auditory instruction sentence in which the disambiguating position word (see Table 1) follows the test word (e.g., a test word plus "near" (**[**kak*ai]) or the same test word plus "above" ([saŋtanɛ])). On *Shape* trials (with the auditory instruction sentence with the name of the shape as the semantic information), two identical printed words are accompanied by different shapes (e.g., 'circle' versus 'triangle'), while the shapes remain in the same positional relation with the word. This time, disambiguation is determined by the name of the shape contained in the instruction sentence that listeners hear. Since the name of the shape occurs later than the position word in the instruction sentence (as shown in Table 1), disambiguation is expected to occur later for *Shape* trials than for *Position* trials.

These semantic trials were included to serve two purposes. First, as Mitterer and McQueen (2009) observed, the inclusion of semantic trials would make it more likely to find phonological effects, presumably by drawing the listener's attention away from the phonological manipulations. Second, these trials allow us to independently determine the critical time window for the *phonological trials*, during which eye-movements are influenced by the position information, but not yet fully disambiguated by the shape information. The critical time window is determined as follows: On *Position* trials, participants should have a preference for the target over the competitor when the position word in the instruction is processed. Because the position words also carry the phonological context information (i.e., the assimilation-triggering context with the /k/-initial word ('near') and the control context with the /s/-initial word ('above')), *Position* trials provide an estimate of the time when the context information starts to be used to guide eye-movements in phonological trials. Similarly, *Shape* trials allow us to estimate when the shape information is processed. Because the shape information also serves a disambiguating function by indicating the target to click on (in both semantic and phonological trials), *Shape* trials can tell us when the phonological context should cease to have an effect on eye-movements.

Method

Participants

Twenty native speakers of Korean, who were undergraduate students at Hanyang University in Seoul, Korea, participated for the study and were paid for their participation.

Materials

Visual stimuli were presented on a computer screen, positioned approximately 60 cm in front of the participants. Each display consisted of four printed words and four geometrical shapes on a screen (see Fig. 2), with the resolution of 1280 by 960 pixels. The bitmap templates for the printed words had a height of 81 pixels and a width of 136 pixels for one-syllable words and 227 pixels for two-syllable words. The geometrical shapes were fitted in a 70×70 pixel square with at least 5 pixels margin. The center of the printed words coincided with the center of the one of the quadrants of the screen, independent of word length. Geometrical shapes were positioned in two fixed locations in relation to printed words (i.e., for "above", shapes were always above printed words, and for "near", they were always on the right of printed words), so that shapes are closer to the center of the screen than printed words as many times as they are farther from the center of the screen than printed words. The positions of the shapes were adjusted to the length of the words, so that the distance between shapes "near" a word and the right-edge of a word was always constant, independent of word length.

We selected 48 minimal pairs (96 words), differing only in the place of articulation of the final consonant, which was either labial or velar. Half of the minimal pairs ended in voiceless stops (/p/ or /k/), the other half of the minimal pairs ended in nasals (/m/ or /ŋ/). A naïve female native speaker of Korean produced these words multiple times

(at least three times) both in the assimilation-triggering context (with the position word /kak*ai/ 'near') and in thee neutral context (with the position word /saŋtan/ 'above), in a natural speaking style that is suitable for connected-speech processes to occur. The most naturally and clearly produced words without any intervening prosodic phrase boundary, confirmed by the two co-authors who were Korean, were then selected to be included in constructing the auditory stimuli. A full list of 48 minimal pairs is given in Appendix 1. It should be noted that the selected words were different in terms of their morphological and semantic status as well as their frequency of occurrence. But as can be seen from quite a large range of stimuli, these factors were randomly distributed between velar-ending and labial-ending conditions, minimizing their possible effects.

The auditory instruction sentences used in the experiment consisted of two cross-spliced parts. The first part contained the start of the sentence and a target word (see Table 1). The second part contained a position word followed by the name of a geometrical shape (which provides the disambiguating information in the phonological trials) and the instruction for a mouse click. Note that Korean has a phonological rule of post-obstruent tensing, by which an obstruent consonant (e.g., /k/ as in /**k**ak*ai/, 'near') following a lenis stop consonant (e.g., /p/ or /k/ in the target words) become a tense (fortis) stop. This means that /kak*ai/, one of the positional words which is at the beginning of the second part of the instruction, sounds different depending on whether it occurs after a stop (i.e., after /p/ or /k/, it becomes [**k***ak*ai],) or after a nasal (i.e., after /m/ or /ŋ/, it is [kak*ai]).[2] Therefore, different renditions of the second part were used to match the phonetic characteristics of the final segment of the preceding word (i.e., stop or nasal). The complete instructions were concatenated offline and presented as single wave files during the experiment.

Apparatus, procedure, and design

The procedure was tailored in such a way that the presentation of a visual display contained no cues to what the target was. Therefore, the randomization procedure balanced the number of times that each quadrant of the screen and each of the four shapes were associated with the target. Moreover, the distractors on a given trial had to be not recognizable as such. To achieve this, we used the same items as targets, competitors and distractors on different trials.

As described above, we also used *semantic trials*, on which the target and the competitor were two identical Hangul character strings, to be distinguished either by the positional relation with the associated shapes (*Position* trials) or by the kind of shapes (*Shape* trials). Distractors on these trials were two tokens of another printed word, so that the appearance of two identical words on the screen did not give that pair away as necessarily containing the target. In addition, distractors were accompanied by the same shape on *position* trials, but in different positional relations, just as the target and its competitor. For example, if the target was under a star and its competitor was near a star, then both of the distractors could have been associated with a circle (one above and one near it). Similarly, on *shape* trials, two identical distractors appeared both under or both near different shapes, just as targets and competitors did. For example, if the target was under a star and the competitor was under a circle, one distractor could be near a triangle and the other near a rectangle.

For the *phonological* trials—the critical trials with minimal pairs such as /tʃik/ and /tʃip/, four different words from two minimal pairs appeared on the screen. As Fig. 2 indicates, one minimal pair ended in nasals, the other in stops. Furthermore, the target and the competitor appeared in the same positional relation with regard to their shapes, and both distractors appeared in the same positional relation to their shapes. That is, if participants were not able to decide what the target was on the basis of the word alone, trials were disambiguated late in the instruction sentence by the shape information.

Participants were tested individually in a sound-damped booth. First, the eye-tracker (an SR Research Eyelink II system, sampling at 500 Hz) was mounted and calibrated. Both eyes were tracked. For the analysis, the eye which produced better results at the validation of the calibration was chosen.

Then, participants saw a written instruction on the screen stating that during the experiment they would hear auditory instructions, presented over headphones, directing them to use the computer's mouse to click on one of the targets in displays that they would see on the computer screen. Note that while the positional word /saŋtan/ is specific, meaning "above", the other positional word /kak*ai/ simply means "near" and hence can refer to any positional relation (i.e., above, below, left, right, etc.). The written instruction therefore stressed that the word "near" was meant to be contrastive to "above", and that a shape "near" a word always appeared on the right side of a word in the visual displays. To make sure that participants got used to such a specific use of the word "near", every participant first completed 10 practice trials. (Note that two of the authors who were native Korean agreed that the use of "near" to mean "on the right side of" was immediately acceptable, and Korean participants also reported that they could immediately associate "near" with "on the right side of" without any hesitation.) The practice trials made use of words that were not the experimental items. After that, 216 main trials were presented, which was composed of 4 blocks of 48 phonological trials (n = 192), plus 24 semantic trials randomly interspersed between the phonological trials. The presentation of the auditory instruction and visual display (printed words and shapes) was controlled with the SR Research program Experiment Builder.

Each participant received a different randomization of trials. Within each block, each of the 48 minimal pair was

[2] The phonological phenomenon that an obstruent becomes a tense obstruent (e.g., /k/ → [k*]) after another obstruent is known as Post-Obstruent Tensing, and it is likely to occur within an intermediate size prosodic phrase in Korean, called the Accentual Phrase, AP (Jun, 1998). So a /k/ + /k/ sequence within an AP becomes a [kk*] sequence. The first homorganic [k] in [kk*] is often considered to be deleted after Post-Obstruent Tensing is applied, resulting in [k*], as a reviewer pointed out. Whether the first [k] in [kk*] is deleted or not, however, does not pose a problem in the present study because /k*/ and /kk*/ are in general pronounced the same. Moreover, a later analysis of the present study suggested that the other half of the critical stimuli with nasals, which therefore were not subject to Post-Obstruent Tensing, showed no significant difference from stimuli with the stops.

used twice, once as a target-competitor pair and once as a distractor pair. If a word in a minimal pair was a target in the first block, the other member of the minimal pair was used as a target in the second block. Both members of the given minimal pair again served as targets in the third and the fourth blocks. With a visual world eye-tracking paradigm, the repetition of a given pair during the experiment is possible, because participants cannot predict what is going to be the target. Nevertheless, the data analysis takes into account that targets were repeated in the second half of the experiment, using the first versus the second half of the experiment as an additional predictor in the analysis of the phonological trials.

In the resulting 192 phonological trials, half ($n = 96$) were filler trials in which the participants heard a labial-ending spoken target word and had to click on the same labial-ending printed target word. The filler trials were created in order to match the number of presentation of labial and velar spoken words. The main experimental trials were the other half ($n = 96$), with velar-ending spoken words. These 96 trials were composed of four conditions of 24 trials, and the conditions were made by crossing two factors: place of articulation in printed target words (labial versus velar) and phonological context (assimilation-triggering versus control). In half of these trials, the printed target word disambiguated by the shape information had a different place of articulation than the spoken word, because the spoken word always ended with a velar. That is, participants hear [. . . tʃʰaŋ . . .], but have to click on the shape paired with the labial-ending printed target word /tʃʰam/. This mismatch should be easier to overcome in the assimilation-triggering context than in the control context, given that the mismatch can be explained by a possible labial-to-velar assimilation (/. . . tʃʰamkak*ai . . . / → [. . . tʃʰaŋkak*ai . . .]).

Most eye-tracking studies using a visual-world paradigm use fixation proportions as the dependent variable. With the current target-shape combinations, it is difficult to determine when a fixation is in fact on a given target. Therefore, we used the Euclidean distance between fixation location and the center of the printed word as the dependent measure (Mitterer, 2011a; Mitterer & McQueen, 2009). (See Appendix 2 for results of analyses using more traditional fixation proportion measures, which provide virtually the same patterns as the one using the Euclidean distance.) In order to get a continuous measure of distance between the fixation and target position, missing data due to blinks and saccades were replaced with the position of the last recorded fixation for blinks, and the upcoming fixation for saccades so that the saccades are grouped with recorded fixations toward the target position (largely in line with McMurray, Clayards, Tanenhaus, & Aslin, 2008). The distances between the eye-gaze and the different shapes were sampled from the ASCI-output of the eye-tracker at, on average, 10 ms intervals from 200 ms before the spoken target word onset to 1.2 s after the onset of the name of a shape. The time line was corrected for the duration of the different parts of the sentences, because critical words (target, position, shape) varied in duration. For instance, the target words had a mean duration of 196 ms. Therefore, 20 samples were taken during the presentation of the target word, with an average inter-sample interval of 9.8 ms (=196 ms/20). A similar correction was applied for the interval between position word onset and shape word onset.

Results

Semantic trials

In the semantic trials, participants mostly clicked on the correct target item, with 6 errors in 480 trials (1.25%). There was no difference in reaction times between the *Position* and the *Shape* trials (2189 vs. 2183 ms). Fig. 3 shows the looks to target, competitor, and pooled distractors for the semantic trials with correct reactions, split by the disambiguation conditions (i.e., Position and Shape). Three thin vertical lines indicate the following three time points in the spoken instruction: the onset of the spoken target word, the onset of the position word (indicating the positional relation between the target and the shape, and providing

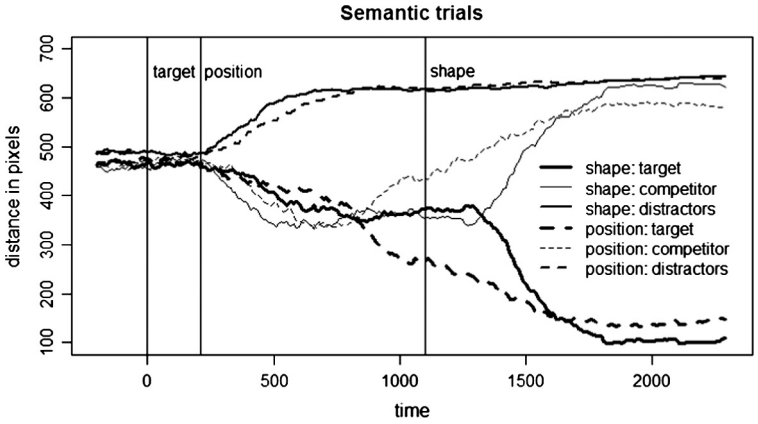

Fig. 3. Distance between fixation position and the different screen objects (target, competitor, pooled distractors, coded with line thickness) for the semantic trials. The data show the typical three-way split in visual-world studies with target, competitors, and distractors. Before target onset, participants have no preference for any of the objects. When they hear the target, this rules out the distractors, and we see a preference of target and competitor over distractors. When the position information disambiguates the target earlier, we see a preference for the target over the competitor arising.

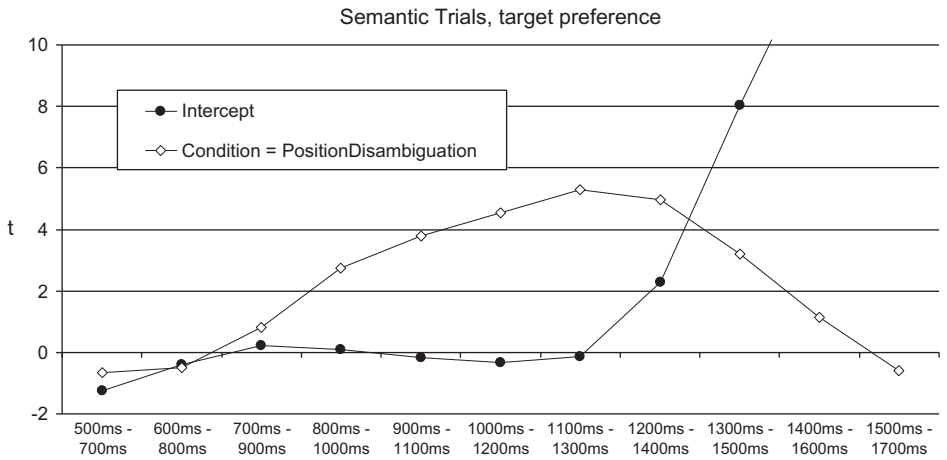

Fig. 4. Results of the sliding window-analyses for the semantic trials. The significant intercept ($t > 2$) indicates that the participants have a preference for the target over the competitor in the object condition. A significant regression weight for the Position Condition indicates that the target preference is larger in this condition than in the Object condition. The last intercept ($t \approx 14$) is not depicted in order to make the critical comparisons more visible.

phonological contexts), and the onset of the shape word. The eye-tracking results show that the participants' looks converged on the target (the thick lines) and away from the competitor (thin lines) faster in the position trials (dashed lines) than in the shape trials (solid lines), reflecting the differential timing of the disambiguating information in the instruction sentence in the two types of trials. Note that, with the Euclidean distance measure, a downwards moving line means more looks towards this objects.

As described above, one purpose of the semantic trials was to estimate when in time participants process the context information and when they have processed the disambiguating shape information. As typical for visual-world eye-tracking studies, we compare looks to targets (objects to be clicked on), competitors (objects that bear some resemblance to the target), and distractors (objects which can easily be ruled out as targets). On semantic trials, the target is the word-shape combination to be clicked on and the competitor is the same word, differing either in accompanying shape or positional relation from the target. The two other words are distractors. We compared the difference between the competitor and target distance between conditions over time. Note that a positive value indicates that the distance between the fixation position and the target is smaller than the distance between fixation position and competitor; that is, participants prefer the target over the competitor. We analyzed this measure (i.e., the difference between the competitor and the target distance) in 11 consecutive 200 ms time windows with a multi-level model with subject and item as random factors and Condition as a fixed effect. Similar approaches have been used to evaluate the time course of the Lateralized Readiness Potential in ERP research (Turennout, Hagoort, & Brown, 1998). The first time window was 500–700 ms and the last 1500–1700 ms after target word onset. The fixed effect of Condition was dummy coded with the shape position mapped on the intercept. That means that a significant positive value for the intercept indicates that participants prefer the target over the competitor in the Shape Condition. The beta weight for Condition = Position indicates whether the preference for the target over the competitor is larger in the Position Condition than in the Shape Condition. It is useful to consider what it means if one of these regression weights (or both) are significantly positive. If only the beta weights for Condition = Position is significant and the intercept is not significantly different from zero, this means that participants are already able to distinguish target and competitor in the Position Condition, but not yet in the Shape Condition. If both are significant, it means that participants are able to distinguish target and competitor in the Shape condition, but are still better in the Position condition than in the Shape condition. If only the intercept is significant and there is no additional effect for Condition = Position, this means that participants are able to distinguish target and competitor in both conditions to the same extent. That is, because the intercept value counts for both condition, an insignificant Condition = Position regression weight does not mean that participants do not distinguish the target and the competitor in this condition. The absence of such an effect only shows that participants do not perform better in the Position condition than in the Shape condition.

The results of this sliding-window analysis are displayed in Fig. 4. In the first four time windows, there is no significant effect, indicating that participants could not distinguish the target from the competitor in any of the conditions. This fits with the descriptive data in Fig. 3, which indicate that a preference for the target over the competitor arises only around 800 ms after target onset, and only in the Position Condition. As Fig. 4 indicates, the effect is significant from 800 ms after target onset until 1400 ms after target onset, while there is no or only a weak effect for the intercept. As indicated above, this means that participants performed better in the Position condition than in the Shape condition. Around 1400 ms, however, the shape information is processed and a significant preference for the target over the competitor in the Shape Condition arises, leading to a significant intercept as well. At the same time, the additional benefit of the earlier disambiguation in

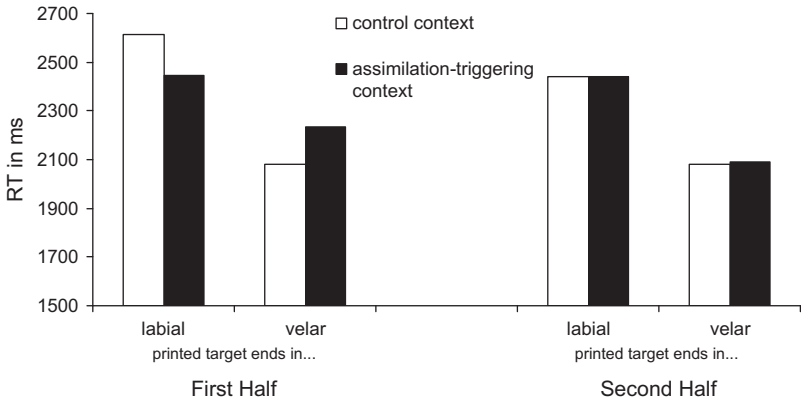

Fig. 5. Reaction times to the critical phonological trials measured from target onset.

the Position Condition disappears. This means that the critical time window, in which the context information has already influenced eye-movements significantly, but the shape information does not yet do so is 800–1400 ms. This window will be used to estimate context effects in the critical phonological trials.

It should be noted that the sliding time window analysis may not be appropriate for certain types of data analysis. Because it involves repeated testings, false positives may arise. Moreover, because it relies on a somewhat arbitrary criterion of statistical significance and is as such sensitive to the sample size, it may not be an ideal way of estimating the exact onset of an effect. Given two lines that start diverging, the onset of a *significant* difference between these lines is likely to be estimated earlier (when the mean difference is still smaller) as the sample size increases and the standard error of the mean difference decreases accordingly. This is in principle undesirable, as estimates of population parameters should be sample-size independent (as the mean, for instance, is). However, the purpose of this analysis was neither to estimate the true onset of an effect nor to show the existence of the semantic disambiguation effect (see Mitterer, 2011a; Mitterer & McQueen, 2009, for independent evidence for such a semantic disambiguation effect). Instead, the present study employed the sliding time window analysis in order to find a most reliable time window during which we can expect significant effects of our experimental variables on other, *unrelated* trials. We therefore believe that the sliding time window analysis was a useful way of determining such an analysis time window.

Phonological trials

In the critical phonological trials, all but one participant mainly accepted the mismatch between spoken and written targets on the vast majority of trials (>20 out of 24), and clicked on the word as indicated by the shape name in the instruction sentence. The accuracy rate was 99.5% in the trials with a velar target (matching the spoken word), and 94.5% and 95% for labial targets (mismatching the spoken word) in assimilation-triggering and control contexts, respectively. Excluding the one participant with more than 50% errors on these trials, the accuracy rates in these conditions were 99.3% (assimilation-triggering context) and 97.8% (control context). However, the error rates were so low that statistical tests (multi-level models with subject and item as random factor and a binomial linking function) showed no significant effect.

The reaction time pattern for the correct trials showed that participants changed how they dealt with the mismatch over the experiment (see Fig. 5). In the first half of the experiment, clicks on printed targets are overall slower if the printed target has a different (i.e., labial) place of articulation than the spoken word, which ended in a velar. Reaction times are, however, especially slow in the control context. Apparently, participants find it easier to accept the mismatch between spoken word and printed target if it can be explained as the result of assimilation. Interestingly, the assimilation-triggering context slows down reactions to velar written targets. This may reflect that a produced velar–velar sequence could be the result of an intended labial–velar sequence, which was subjected to assimilation. In the second half of the experiment, when items are repeated as targets, there only remains an effect of mismatch, with slower responses for mismatches between spoken and written targets. These patterns were statistically significant, as a multi-level mixed effect model for the reaction times as dependent variable with Context (assimilation triggering vs. control), Target Place (labial vs. velar), and Experiment Part (1st half vs. 2nd half) showed. We first fitted a model with all interactions between the main effects, and one with all but the three-way interaction. A model comparison revealed a significant three-way interaction, as the model with the three-way interaction was found superior in a model comparison ($\chi^2_{(df=3)} = 19.2$, $p < 0.001$). In order to understand the nature of this interaction, we analyzed the data separately for the first and the second half, using the remaining two fixed effects—Target Place and Context—and their interaction. For each half, we compared a model with and without an interaction between the two fixed effects. This revealed a significant interaction for the first half ($\chi^2_{(df=1)} = 28.7$, $p < 0.001$) but not for the second half ($\chi^2_{(df=3)} = 0.7$, $p > 0.1$). Therefore, another model was run without the interaction for the data from the second half. Note that such pruning of interactions is critical, as interaction and main effects are not necessarily linearly independent in the regression approach applied in a linear mixed-effect

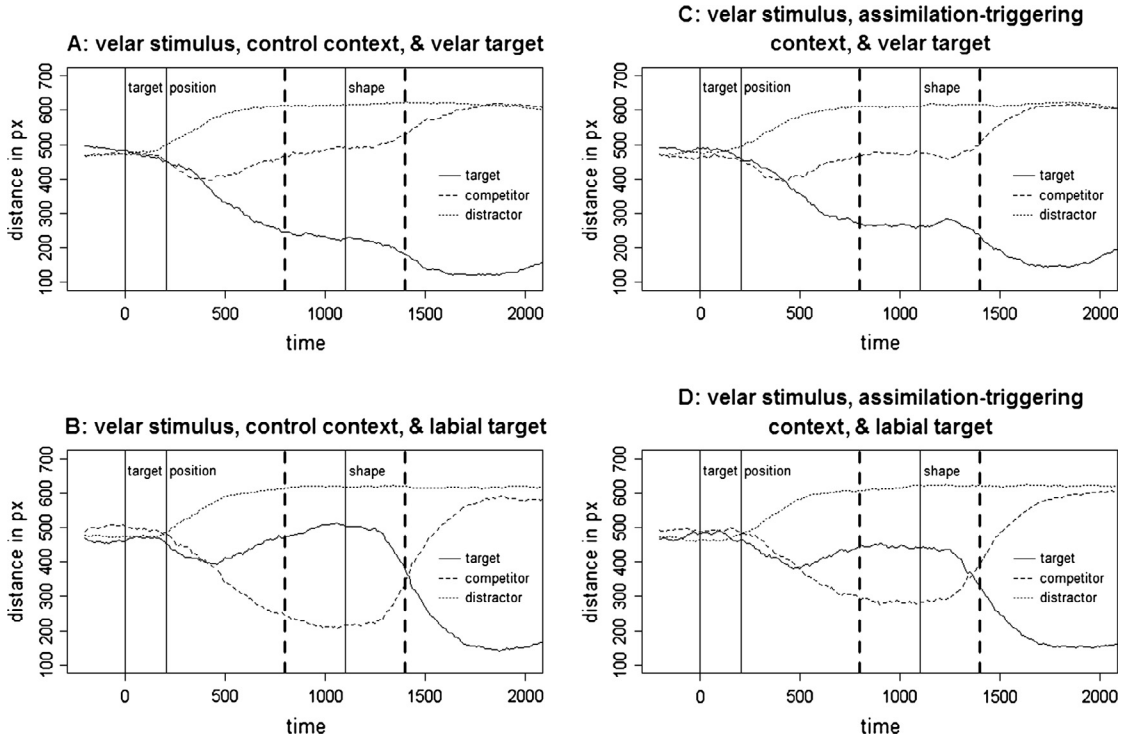

Fig. 6. Distance between fixation position and the different screen objects (target, competitor, pooled distractors, coded with line thickness) for the experimental phonological trials. The thin lines indicate the onset of target, position, and object word in the sentence. The thick dashed lines indicate the critical time window (800–1400 ms) as determined from the semantic trials, on which the data analysis is based.

model. The model with only the main effects and no interaction for the second half showed no effect of Context ($b = 20.1$, $p_{MCMC} > 0.1$) but an effect of Target Place for the data of the second half ($b = -347$, $p_{MCMC} = 0.0001$), with faster reactions to velar targets.

For the first half, we further broke down the interaction between Target Place and Context and investigated the effect of the Context separately for the labial and velar targets. This reveals that the assimilation-triggering context speeded up responses if the printed target had a mismatching labial place ($b = -160$, $p_{MCMC} = 0.0002$) but slowed down responses when the target had a matching place of articulation ($b = 141$, $p_{MCMC} = 0.0004$).

The eye-tracking data for these trials are depicted in Fig. 6, again comparing looks to the eventual target, its competitor (the other member of the minimal pair) and the two distractors. Note that the two distractors always differed in manner of articulation of the final consonant from the target. Hence, the distractors were easy to distinguish from the target. As the data shows, participants were mostly influenced by the surface place of articulation of the spoken word. That is, when they heard a velar-ending word, they looked at the corresponding printed word. This is the target on trials on which the shape information indicates to click on the velar-ending word (top panels of Fig. 6), but it is the competitor on trials on which the shape information indicates to click on the labial-ending word (lower panels of Fig. 6).

The overall effect was, however, moderated by the phonological context. In the assimilation-triggering context (the two right panels of Fig. 6), participants had less of a preference for the velar printed word than in the control context (the two left panels of Fig. 6). To statistically confirm this pattern, we analyzed the pattern of target preferences (that is, competitor distance minus target distance) using the predictors Target Place and Context and their interaction. The target preference was calculated for the time window 800–1400 ms after target onset, the time during which the context is already processed but the object information has not completely disambiguated the target yet, as estimated from the results obtained in the semantic trials. The analysis revealed a main effect of Target Place ($b_{\text{TargetPlace = Velar}} = 517$, $p_{MCMC} = 0.0001$), no overall effect of context ($b_{\text{Context = Assimilation-triggering}} = 31$, $p_{MCMC} > 0.1$) and an interaction of Target Place and Context ($b_{\text{Context = Assimilation-triggering} \times \text{TargetPlace = Velar}} = -180$, $p_{MCMC} = 0.0001$). The negative beta weight for the interaction indicates that the preference for the velar-bearing target (a positive effect, +517 pixels) was less significantly reduced (by 180 pixels) in the assimilation-triggering context. (See Appendix 2 for similar results from more traditional analyses using fixation proportion measures.)

Additional models tested whether the interaction of Target Place and Context was modified by additional variables, such as manner of articulation of the final consonant (stop or nasal) or with experiment half. Models comparisons of the simpler model with models including such additional terms did not improve the model's fit significantly (manner of articulation: $\chi^2_{(df=4)} = 3.3$, $p > 0.1$, experiment half:

$\chi^2_{(df=4)} = 1.2$, $p > 0.1$). That is, the pattern of eye-movements was, in contrast to the reaction time results, stable over the course of the experiment.

Discussion

The purpose of this experiment was to test how Korean listeners deal with the consequences of labial-to-velar place assimilation. The question was whether the ensuing mismatch between spoken and printed target word resulting from assimilation would be treated as any mispronunciation, or whether listeners would take into account that the context could trigger an assimilation. To this end, we compared trials in which participants had to click on a labial-ending printed target while hearing a velar-ending spoken word in two conditions. In one condition, the target word was followed by the velar consonant /k/, which is a context that triggers labial-to-velar assimilation. In the other condition, the target word was followed by the alveolar consonant /s/, which was a control context in which no assimilation could occur. The context clearly influenced how listeners overcame the mismatch. Both the reaction time data as well as the eye-tracking data indicated that listeners found it easier to overcome the mismatch in the assimilation-triggering than in the control context.

At this point, it is worth discussing, as suggested by a reviewer, whether the results may be due to perceptual learning (Mitterer, Chen, & Zhou, 2011; Norris, McQueen, & Cutler, 2003): Because participants had to click on printed words ending in a labial even when they heard a physical velar sound, they might have been habituated to click on a labial-ending word when they heard a velar sound. The change in reaction time patterns over the course of the experiment could be seen as evidence for such an account. However, if perceptual learning were the only cause that biased listeners to accept any velar as a possible labial, it is hard to explain why perceptual learning would apply only to the assimilation triggering context. In our experiment, the mismatch occurred in both assimilation-triggering and control contexts; therefore, perceptual learning should have been observed in both contexts as well. It was, however, clearly not the case: the acceptance of velars as labials was observed only in the assimilation-triggering context. It is, therefore, unlikely that the critical findings of the present study are due to perceptual learning. Moreover, perceptual learning seems to take place only after a substantial number of repetitions of an unusual pattern (e.g., Poellmann, McQueen, & Mitterer, 2011) suggested that at least 10 repetitions of the unusual item would be needed for perceptual learning to take place). It is then expected that the acceptance of velars as labials should increase over the course of the experiment. However, as reported above, the eye-tracking results showed no such trend, being statistically equivalent for the first and second half of the experiment.

As the result cannot be explained by such an alternative interpretation, it raises two theoretical questions. First, Gow and Im (2004) showed that detection of velar sounds /k, g/ was not facilitated after the assimilated [ŋ] (derived from /m/) than after the underlying /ŋ/, evident in both Korean and English listeners' performances. They then argued that both Korean and English listeners are not able to undo the consequences of labial-to-velar assimilation by parsing features, because the assimilation is complete and therefore there is no phonetic residue of the underlying phoneme that listeners can make use of. While Gow and Im showed no effect of assimilation on the post-assimilation context word, no regressive context effect should naturally follow in line with the predictions by the feature parsing theory (Gow, 2003), although the regressive context effect was not tested in Gow and Im. The results of Experiment 1, however, showed that listeners made use of the context information of complete labial-to-velar assimilation, even though there was no acoustic–phonetic residuals in the speech signal. This undermines the theoretical assumptions of the feature parsing account.

Moreover, Gow and Im (2004) claimed that language-specific knowledge did little to help listeners to compensate for this assimilation. This claim was based on the finding that native and non-native listeners (i.e., Korean and English listeners) did not differ in their perception of complete labial-to-velar assimilation. That is, according to the data of Gow and Im, native and non-native listeners are equally *inept* in dealing with this form of assimilation. The current results, however, indicate that native listeners are not at all inept in dealing with this form of assimilation. But with the data with Korean listeners alone, one cannot make a decisive conclusion about whether this context effect is a consequence of Korean listeners' exposure to this assimilation rule in Korean.

Note that other accounts for compensation of assimilation make different predictions here. The phonological-inference account (e.g., Gaskell, 2003) argues that this context effect is due to language-specific phonological knowledge, while the perceptual-integration account (Mitterer, Csépe, & Blomert, 2006) assumes that low-level perceptual processing based on general auditory mechanism is at least partly responsible. In order to test the role of language-specific phonological knowledge, we would need to run the same experiment with listeners who are not familiar with the Korean language or with any other language with the same assimilation rule. While finding a listener group that can satisfy such a condition is not a problem, the problem lies in the fact that non-native Koreans simply cannot perform the same experimental task because the experiment was geared to testing Korean listeners' performance with native spoken and written test materials. We therefore decided to use the format of a categorization task as categorization tasks can reveal context effects in speech perception even if the participants—be they human (Mann, 1986; Viswanathan, Magnuson, & Fowler, 2010) or non-human (Lotto, Kluender, & Holt, 1997)—are not familiar with the language the stimulus material is based on.

However, before we can test non-native perception of Korean labial-to-velar place assimilation with a categorization task, we first need to test whether the phonological context effect found with Korean listeners in the eye-tracking task can also be found with Korean listeners in a categorization task. Note that the context effect in the perception of /t/-reduction in Dutch was only found with an eye-tracking task (Mitterer & McQueen, 2009), but not in a categorization task (Mitterer & Ernestus, 2006). There is thus a possibility that even Korean listeners, who showed the context effect in the

eye-tracking experiment, do not show a phonological context effect in a categorization task. In such a case, we would not need to proceed with non-native listeners.

Such a result would, nevertheless, be informative, in a sense that it would show a dissociation of the context effect over tasks. Such dissociation would be easier to explain in the framework of the phonological inference account than in the framework of the perceptual-integration account. The phonological inference account assumes that the assimilated sequence is first perceived *as is* at a phonetic level and is then fed to the next stage in which a phonological filtering applies. Given the assumption that the phonetic categorization occurs at the phonetic level, the failure to find the context effect with the phonetic categorization task could be interpreted as supporting the dissociation between the phonetic and the phonological filtering stages as assumed by the phonological inference account. The perceptual-integration account, on the other hand, assumes that context effects are a consequence of early and mandatory auditory processes (somewhat in line with proposals form an auditory framework for speech perception, see Holt & Lotto, 2008), and therefore predicts that the context effect should be evident in any task (see Mitterer, 2011b, for a fuller development of this argument). Finding a context effect, however, does not allow us to decide between the two accounts. Instead, such a finding relegates the decision between these two accounts to the question of whether the context effect can also be found with listeners that do not have any language-specific phonological knowledge about labial-to-velar place assimilation.

Experiment 2 hence tries to replicate the context effect found in Experiment 1 with a categorization task with Korean listeners, which will serve two purposes—i.e., determining whether or not we should run the identical experiment with listeners that are not familiar with Korean and whether the dissociation between the auditory and the phonological processing could arise as predicted by the phonological inference account.

Experiment 2

To implement a categorization task, we selected four one-syllable minimal pairs (see Table 2) which showed a clear context effect in Experiment 1. From the natural endpoints, we generated a six-step continuum using a morphing technique. We then presented these in the context of one syllable, for instance [tʃiŋsaŋ] or [tʃiŋka], and asked participants whether they had heard /tʃiŋ/ or /tʃim/ (see below for details). The critical question was whether the context effect observed in Experiment 1 would also be observed in this task, resulting in more clicks on the labial word if the context is velar.

Table 2
Stimuli used in Experiments 2 and 3.

Korean Minimal Pair in IPA	Dutch transcription	Dutch transcription in IPA
[tʃʌm]—[tʃʌŋ]	tsom – tsong	[tsɔm]—[tsɔŋ]
[tʃim]—[tʃiŋ]	tsim – tsing	[tsɪm]—[tsɪŋ]
[tʃip]—[tʃik]	tsip – tsik	[tsɪp]—[tsɪk]
[jʌp]—[jʌk]	jop – jok	[jɔp]—[jɔk]

Method

Participants

Twenty native speakers of Korean participated in the study. The study was conducted at Hanyang University in Seoul, Korea, and participants were recruited from the student body of the University and were paid for their participation.

Materials

Four minimal pairs were selected from the 48 used in Experiment 1. The criteria were as follows: The minimal pairs should be the ones that showed a substantial context effect, with remarkable differences in numerical values, in Experiment 1, should consist of only one syllable to facilitate testing with non-native listeners, and none of the members of the pair should be an offensive word. The selected pairs are displayed in Table 2.

For each pair, four intermediate steps were generated using the STRAIGHT audio morphing algorithm (Kawahara, Masuda-Katsuse, & de Cheveigné, 1999). This algorithm decomposes the speech signal into three parameters: a voice source, a noise source and a dynamic spectral filter with time windows of 10 ms. Interpolation is achieved by mixing the parameters first, and then generating a signal out of these mixtures again. The time-aligned version also mixes the parameters relative to the supplied time anchors, that is, if a vowel were 80 ms long in one utterance and 100 ms in another, the resynthesized vowel would be 90 ms long and the parameters used for the vowel midpoint would come from the fourth 10 ms time window of the shorter sound and the fifth time window from the longer sound. Six stimuli were created by mixing the stimuli in different proportions with a step of 20%. The resulting stimuli hence range from [0% velar coda + 100% labial coda] to [100% velar coda + 0% labial coda] in steps of 20%. Each of the resulting stimuli was concatenated with the first syllable of both the assimilation-triggering and control context word (i.e., [ka] and [saŋ]) to result in 48 stimuli (4 minimal pairs $*$ 6 version $*$ 2 contexts).

Apparatus, procedure, and design

Experiments were run with a standard PC and stimulus presentation was controlled using MatLab using the PsychToolBox (Brainard, 1997; Kleiner, Brainard, & Pelli, 2007). In order to generate an experiment that could also be performed with speakers that do not know Korean, the minimal pairs as well as the context were blocked, so that the variation from trial to trial was minimal. That is, participants would hear a long series of stimuli with the structure [tʃiCsaŋ], in which C is a member of the [ŋ]–[m] continuum. This allows the listeners to focus on the minimal differences between the stimuli. The order of the four minimal pairs, as well as the order of the context conditions, was counterbalanced over participants using a Latin-Square design.

Participants received all instructions via the screen. The first introduction told them that they would see two Korean words on a screen and hear a Korean phrase starting with one of these words. The instruction screen gave a general introduction for the categorization task and stated that participants would have to decide which of the two printed

words presented on the screen fitted better with what they had heard, by clicking on one of the words. Three additional instruction screens appeared, leading the participants to go through a three-step procedure (familiarization, categorization of endpoints in isolation, and categorization in context).

The first of the three instruction screens stated that once the participants click a mouse, they would hear examples of the two words. Upon clicking, words with the natural endpoints (i.e., the most extreme versions) of a continuum were played four times without context, with a 700 ms inter-stimulus interval. In sync with the audio, the screen switched between the Korean transcriptions of the two words. This familiarization procedure was probably superfluous for the native speakers of Korean that participated in this experiment, but was implemented anyway so that both Koreans and non-Korean participants (in Experiment 3) could go through the same procedure. After hearing the endpoint examples, the second instruction screen stated that participants would now hear these words in isolation and that they have to decide which word they have heard by clicking on one of the transcriptions displayed on the screen. When participants clicked on a mouse button, this instruction screen disappeared and participants heard a series of 10 stimuli (each endpoint stimulus $\times$ 5 repetitions) which they had to categorize without a context. After this training session, a final instruction screen appeared, telling the participants that they would now hear the words in a context, but that the context is to be ignored. This three-step procedure (familiarization, categorization of endpoints in isolation, and categorization in context) was repeated every time when a new stimulus pair was introduced.

Within the block for a given syllable, the context condition was also blocked. Each participant hence categorized each of the six stimuli of one minimal pair in one context (assimilation-triggering or control) six times before moving onto another set of stimuli. Within the set of 36 stimuli, stimuli were presented as six permutated series of the six base stimuli. The experiment consisted of 40 practice trial and 288 experimental trials (4 syllables $\times$ 2 contexts $\times$ 36 trials per block).

The results were analyzed using linear mixed-effect models with a binomial linking function to account for the categorical nature of the response variable. The fixed effects were manner of articulation (oral or nasal), Level on the velar–labial continuum (6 steps), and context (assimilation-triggering or control). Participant was added as a random factor. Even though Level is in principle a continuous predictor, it was entered as a factor to allow for non-linearity. The analysis started with a full model and insignificant interactions were pruned. This pruning is crucial when using linear-mixed effect models with the (standard) treatment coding of the variables in linear-mixed effect models.

Results

The results from the training phase indicated that the participants found it easier to hear the difference between the velar and the labial endpoints for nasals (94% correct) than for stops (82% correct). There was an asymmetry for the stop stimuli so that velar stimuli were categorized as labial more often (31%) than vice versa (6%). This difference between stops and nasals is also evident in the results of the main experimental trials displayed in Fig. 7, with steeper identification functions for the nasal stimuli than for the stop stimuli. Nevertheless, for both types of stimuli, we see a context effect emerging at the "velar" end of the continuum. Here, velar–velar combinations lead to more labial responses than velar–alveolar combinations, as shown by straight lines (velar–velar, assimilation-triggering context) being lower in Fig. 7 than dotted lines (velar–alveolar, control context). This context effect is also evident in the statistical analysis. The full model was pruned of the three-way interaction between all factors and the context-by-manner interaction. The final model is displayed in Table 3. The Intercept value indicates the likelihood of a velar response for the cell of the design mapped onto the intercept, and the other regression weights the changes from the intercept in logistic space ($\text{logit}(p) = \log[p/(1-p)]$). Note that in logistic space, the chance performance of 50% is mapped on zero ($\log[0.5/(1-0.5)] = \log(1) = 0$). The significant negative intercept hence indicates that the participants responded mostly chose the labial response in the intercept condition (cf. Fig. 7). Unsurprisingly, there are strong effects for the continuum steps, with more velar responses the more velar the stimulus. The insignificant main effect of context shows that, at the intercept level, there is no difference between the two context conditions. However, a significant context effect arises at the "velar" end of the continuum, where

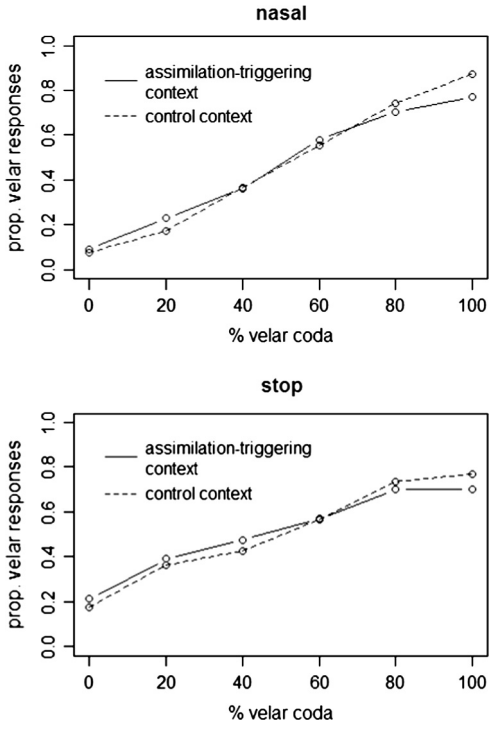

Fig. 7. Proportion of velar responses in Experiment 2 by Korean listeners. The top panel shows the results for nasals, the bottom panel the result for stops. The context effect at the velar end of the continuum (100% velar coda) is significant.

Table 3
Final model for the data of Experiment 2 (Korean participants).

Source	Regression weight	Estimate (SE)	*z*	*p*
Intercept	Context = Control, Manner = Nasal, Contimuum Step = 1	−2.48 (0.19)	−12.75	<0.001
Continuum	ContinuumStep = 2	1.05 (0.2)	5.17	<0.001
	ContinuumStep = 3	1.9 (0.19)	9.83	<0.001
	ContinuumStep = 4	2.76 (0.19)	14.36	<0.001
	ContinuumStep = 5	3.48 (0.2)	17.63	<0.001
	ContinuumStep = 6	4.1 (0.21)	19.69	<0.001
Context	Context = assim.	0.24 (0.19)	1.24	0.21
Manner	Manner = Stop	0.99 (0.2)	4.88	<0.001
Continuum × context	ContinuumStep = 2:Context = assim.	−0.02 (0.24)	−0.09	0.93
	ContinuumStep = 3:Context = assim.	−0.14 (0.23)	−0.59	0.55
	ContinuumStep = 4:Context = assim.	−0.18 (0.23)	−0.76	0.44
	ContinuumStep = 5:Context = assim.	−0.42 (0.24)	−1.74	0.08
	ContinuumStep = 6:Context = assim.	−0.78 (0.25)	−3.09	<0.01
Continuum × manner	ContinuumStep = 2:Manner = Stop	−0.09 (0.25)	−0.36	0.72
	ContinuumStep = 3:Manner = Stop	−0.62 (0.24)	−2.56	<0.05
	ContinuumStep = 4:Manner = Stop	−0.99 (0.24)	−4.05	<0.001
	ContinuumStep = 5:Manner = Stop	−1.03 (0.25)	−4.09	<0.001
	ContinuumStep = 6:Manner = Stop	−1.53 (0.26)	−5.88	<0.001

an assimilation-triggering context significantly lowers the likelihood of a velar response. The main effect of Manner and its interaction with the continuum can all be explained by the shallower slope of the categorization function for stops. At the intercept level, where the stimulus is labial, participants give more "velar" responses to stop than to nasal stimuli. This reverses at the end of the continuum, where participants give more "labial" responses to stop than to nasal stimuli.

Discussion

The results of this experiment converge with those of Experiment 1. When Korean listeners hear a sequence of two velar consonants, they take into account the possibility that the sequence might have arisen from an underlying labial–velar sequence that has undergone place assimilation. This makes them more likely to categorize a velar stimulus as labial in the velar context. The effect was robust over manners of articulation for the test sounds. The context effect can be said to be a 'contrastive' phonological context effect in the sense that the velar-like sound before a velar in the context is likely to be perceived as a labial which is indeed phonologically contrastive with the velar.

It is important to note that this context effect was significant only for the velar endpoint of the 6-step labial-to-velar continuum. At all other steps, context did not influence the categorization performance significantly. These perception results appear to mirror what happens in production. The acoustic data measured by Gow and Im (2004) and the articulatory data in Son et al. (2007) both showed that Korean labial-to-velar assimilation is, if it occurs, complete. An examination of the assimilation data in casual speech provided by Jun (1996) also revealed that more than 80% of the assimilated tokens were complete, although some assimilated tokens were incomplete. While there is some discrepancy between the two studies, it is clear that Korean labial-to-velar assimilation is most likely to be complete. Korean listeners' phonological knowledge could be based on the frequent 'complete' patterns of labial-to-velar assimilation in Korean. This may account for why the context effect in the phonetic categorization was found only with the most clear velar stimuli which matched the completely assimilated velar sounds that they had to deal with in processing labial-to-velar assimilation in real life.

However, an alternative interpretation should be considered as suggested by a reviewer. That is, the context effect could also be considered in terms of differences in slope as a shallower slope was found in the assimilation triggering context. When the continuum was used as a numeric factor,[3] the continuum by context interaction turned out to be significant ($b = -0.074$, $z = -3.78$, $p < 0.001$), giving rise to a shallower slope in the assimilation-triggering context. This is in fact another way to conceptualize compensation for assimilation. This is in line with the findings by Mitterer, Csépe, and Blomert (2006), who observed that the identification function slope for a continuum between an underlying and potentially assimilated segment tends to be shallower in the context that can trigger the assimilation in question. For example, a continuum between /n/ and /m/ was found to be shallower in bilabial context that can trigger an assimilation of an underlying /n/ to an [m]. This shallower slope can be interpreted functionally: In the assimilation-triggering context, listeners are less certain about the interpretation of the input signal, which, in turn, minimizes the potential mismatch between the perceived utterance and the canonical form. We therefore propose that a reduction in slope is an alternative characterization of the compensation for assimilation effect.

The results of both experiments therefore provide two independent lines of evidence that there is a contrastive phonological context effect in the perception of Korean labial-to-velar place assimilation. It cannot be, nevertheless, assured that the contrastive context effect has stemmed strictly from

[3] It is noteworthy that the model with continuum as a categorical predictor yields a better model fit ($X(12) = 25.56$, $p < 0.05$).

language-specific phonological knowledge of labial-to-velar place assimilation until non-native listeners with no experience with Korean are tested with the same experimental materials. Experiment 3 therefore investigates this question by replicating the experiment with Dutch listeners who are not familiar with Korean and do not have prior experience with labial-to-velar assimilation in their native language, either.

Experiment 3

The procedure of Experiment 2 was already tailored to the needs of non-native listeners, with stimulus specific familiarization phases. However, the response options were still presented in Korean script in Experiment 2. This needed to be adapted for Dutch listeners. Therefore, three naive speakers of Dutch were asked informally to transcribe the stimuli into Dutch script. To this end, they could hear the stimuli as often as they wanted. Table 2 shows the test words spelled in Dutch agreed by these transcribers, which were then used as response options in the 2AFC task.

Method

Participants

Twenty-two native speakers of Dutch participated in the study. The participants were students from the Radboud University Nijmegen. None of the participants were familiar with Korean. Participants were financially compensated.

Materials, apparatus, procedure, and design

Materials, Apparatus, Procedure, and Design were the same as in Experiment 2, apart form the instructions and the display for the alternatives in the 2AFC task. The initial instruction stated that the stimuli were Korean and that identification might be difficult. Otherwise, the instructions were translations of the instructions used in Experiment 2.

Results

With regard to the training phase, Dutch participants unsurprisingly performed less well than the Korean participants, with overall 70% correct identifications. They showed similar performance for nasals and stops (70% and 68.6% correct respectively). For the main analysis, we rejected the data from two participants, who were not able to distinguish the stimuli in the training phase.

Fig. 8 shows the data from the 20 participants that did perform above chance in the training phase. Table 4 shows the results of the final regression model after pruning of insignificant interactions. The final model shows the following effects. Dutch participants are able to use the cues to coda identity, that is, more velar responses were given to stimuli that contained more cues to a velar. This effect is smaller for stops than for nasals. That is, the negative coefficients for ContinuumStep = [4,5,6] × Manner = Stop are negative, meaning that the difference in proportion of velar responses from the first continuum are smaller if the trials with stops as targets. A separate analysis with only the stop data, however, showed that the continuum also influenced

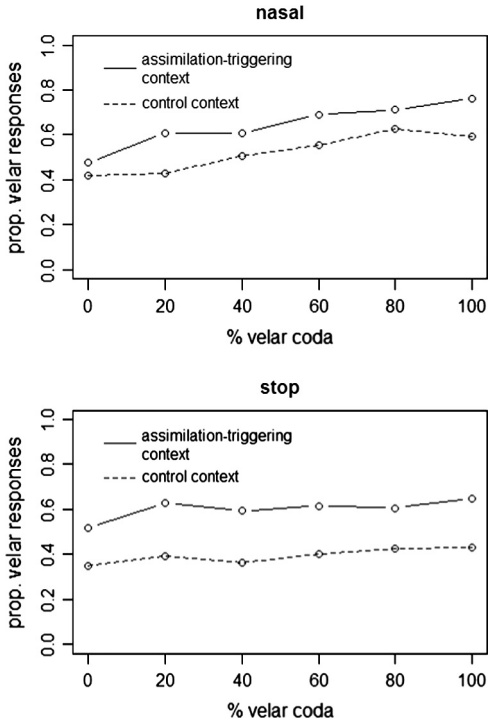

Fig. 8. Proportion of velar responses in Experiment 3 by Duch listeners. The top panel shows the results for nasals, the bottom panel the result for stops.

the categorization of stops. There also is an integrative context effect, meaning that *more* velar responses were given if the following context was velar (i.e., assimilation-triggering). This is an effect in the opposite direction of the contrastive context effect in the Korean data from Experiment 2. This integrative effect is larger for stops and enhanced at the velar end of the continuum (see the regression weights for Context × Manner and Context × Continuum).

As the context effect for the Korean participants might also be characterized as a shallower slope in the assimilation triggering context, we also analyzed Dutch listeners' data with Continuum as a numeric variable. This analysis revealed no significant interaction of Context and Continuum ($b = 0.017$, $z = 1.028$, $p > 0.2$), even though the overall effect of continuum was highly significant ($b = 0.1$, $z = 8.84$, $p < 0.001$).

Discussion

The Dutch participants show not only shallower identification functions than the Korean participants, but they also show lower "correct" responses than the Korean participants even at disambiguating endpoints, although exactly the same speech materials were used. This discrepancy between the two listener groups can be accounted for by differential phonetic implementation of phonological contrasts, depending on phonological structure of listeners' native language. Korean has a phonological process in which stops in the word-final (coda) position are obligatorily unreleased, while Dutch stops are generally released.

Table 4
Final model for the data of Experiment 3 (Dutch participants).

Source	Regression weight	Estimate (SE)	*z*	*p*
Intercept	Context = Control, Manner = Nasal, Contimuum Step = 1	−0.21 (0.11)	−1.92	0.06
Continuum	ContinuumStep = 2	0.29 (0.13)	2.17	<0.05
	ContinuumStep = 3	0.44 (0.13)	3.31	<0.001
	ContinuumStep = 4	0.72 (0.13)	5.41	<0.001
	ContinuumStep = 5	0.93 (0.14)	6.87	<0.001
	ContinuumStep = 6	0.97 (0.14)	7.15	<0.001
Context	Context = assim.	0.3 (0.14)	2.07	<0.05
Manner	Manner = Stop	−0.06 (0.13)	−0.47	0.64
Context × manner	Context = assim. × Manner = Stop	0.33 (0.11)	3.00	<0.01
Continuum × context	ContinuumStep = 2:Context = assim.	0.39 (0.19)	2.08	<0.05
	ContinuumStep = 3:Context = assim.	0.22 (0.19)	1.18	0.24
	ContinuumStep = 4:Context = assim.	0.27 (0.19)	1.43	0.15
	ContinuumStep = 5:Context = assim.	0.11 (0.19)	0.58	0.56
	ContinuumStep = 6:Context = assim.	0.38 (0.19)	1.99	<0.05
Continuum × manner	ContinuumStep = 2:Manner = Stop	0.04 (0.19)	0.20	0.84
	ContinuumStep = 3:Manner = Stop	−0.25 (0.19)	−1.32	0.19
	ContinuumStep = 4:Manner = Stop	−0.41 (0.19)	−2.16	<0.05
	ContinuumStep = 5:Manner = Stop	−0.58 (0.19)	−3.05	<0.01
	ContinuumStep = 6:Manner = Stop	−0.52 (0.19)	−2.72	<0.01

Cho and McQueen (2006) indeed showed that Korean listeners were better in identifying unreleased coda stops than Dutch listeners, suggesting that Korean listeners make a better use of formant transitions in the preceding vowel in identifying the following consonant. One might, however, argue that the poorer performance by Dutch listeners (compared to Korean listeners) for the nasals may not be straightforwardly interpretable, given that similar nasals occur in both languages. However, this can also be accounted for by the substantial mismatch in vowels between the Korean form and the Dutch transcription as shown in Table 2. The Korean vowels used in the experiment were qualitatively different from Dutch vowels, and therefore could be difficult to interpret for Dutch listeners, as also discussed in Cho and McQueen (2006). This may explain the poorer performance on the nasals as well as on the stops. In particular, given that they had to process the unfamiliar non-native (Korean) vowels and given that speech stimuli were based on casually (rather than carefully) spoken speech materials, Dutch listeners may have found it more difficult to interpret the formant transitions into the nasals.

The most important fact is that in processing exactly the same speech materials as used with Korean participants, the Dutch participants not only showed a different pattern in the identification functions, but they also revealed a completely different context effect. Recall that Korean listeners showed a contrastive phonological context effect for the most clear velar stimuli in the labial–velar continuum–i.e., a velar-like sound before another velar in the context is likely to be perceived as a labial. In contrast, Dutch listeners perceived an ambiguous sound more likely as a velar sound if it was followed by another velar. That is, Dutch listeners showed an 'integrative' effect for all stimuli across the continuum integrating their perception of an ambiguous sound with the perception of the following velar sound. As discussed in the Discussion of Experiment 2, the context effect observed with Korean listeners may also be characterized as the difference in identification function slope between the assimilation-triggering vs. the control contexts. Dutch listeners, however, fail to show compensation for assimilation in terms of the slope reduction, as their slopes are not statistically different between the assimilation-triggering and the control contexts.

Alternatively, however, this effect, especially for the stop case, may be interpreted as coming simply from the fact that the first member of the stop–stop sequence is not related in connected speech in Dutch. Dutch participants therefore may have heard the sequence not as a geminate, but rather as a single stop. The release of the second velar (i.e., the velar in the following context) could then be used as a critical determinant for the place of articulation of the test consonant, yielding the observed pattern of results. This explanation, however, does not work for the nasals. Although the nasal target had separate cues to its place, a similar perceptual pattern was still observed, as was the case with stops. It is conceivable, however, that the difficulties with the stop stimuli may also affect the perception of the nasal stimuli, which are, *a priori*, easier to deal with for Dutch listeners (see, e.g., Van der Heijden, Hagenaar, & Bloem, 1984, for an example of such cross-trial transfer of difficult trials).

In Experiment 4, we therefore continue to explore the perception of velar–velar sequences, this time by testing yet another listener group, native listeners of American English, with the same test materials as used with Korean and Dutch listeners in phonetic categorization tasks. In English, an unreleased stop before another stop occurs quite commonly, and other types of place assimilation (e.g., coronal-to-velar or coronal-to-labial) occur more robustly than in Dutch (cf. Ernestus, 2000). Thus, despite the cross-linguistic difference between English and Dutch, if English listeners show a similar integrative effect as Dutch listeners, a stronger conclusion could be made that the contrastive phonological context effect is driven by language-specific phonological knowledge, while the integrative effect arises when listeners have no such language-specific knowledge.

Experiment 4

Method

Participants

Twenty-two native speakers of American English participated in the study. They were either visiting Korea for a short time or had lived in Korean for some time, but none of them had learned Korean. They were paid for participation.

Materials, apparatus, procedure, and design

All other experimental aspects were the same as in Experiments 2 and 3, except for the instructions and the display for the alternatives in the 2AFC task, which were written in English. The initial instruction stated that the stimuli were Korean and that identification might be difficult. Otherwise, the instructions were translations of the instructions used in Experiment 2.

Results and discussion

With regard to the training phase, English participants answered correctly on 80% of the trials, which is less well than the Korean participants (88% correct) but slightly better than the Dutch participants (70% correct). They showed similar performance for nasals and stops (82% and 78% correct respectively) as was with Dutch listeners. All participants were able to distinguish the stimuli in the training phase, so all 20 participants were kept for the main analysis.

Fig. 9 shows the data from the main experimental trials. The results of the final regression model after pruning of insignificant interactions are summarized in Table 5. English participants are able to use the cues to coda identity, and give more velar responses if the stimulus contains stronger cues for a velar. This effect is smaller for stops than for nasals, just as for Dutch participants. That is, the negative coefficients for ContinuumStep = $[3 \cdots 6] \times$ Manner = Stop mean that the difference in proportion of velar responses from the first continuum step are smaller on the trials with stop targets than with nasal targets. A separate analysis with only the stop data, however, showed that the continuum also influenced the categorization of stops. Again as was the case with the Dutch listeners, there also is an integrative context effect, contrasting with the contrastive context effect in the Korean data from Experiment 2. That is, English listeners give more velar responses if the context is assimilation-triggering (i.e., velar).

All in all, the data are very similar to those obtained from the Dutch listeners. This is true even though English is different from Dutch in that English stops are often unreleased when followed by another stop. Furthermore, English listeners who participated in the present study had some degree of exposure to Korean. Nevertheless, they performed similarly as the Dutch listeners did. This lends support to our interpretation that the contrastive effect observed with Korean listeners is due to their language-specific knowledge. The integrative effect observed with Dutch and English listeners is similar to integrative effects found in compensation for coarticulation studies (Fowler, 2006):

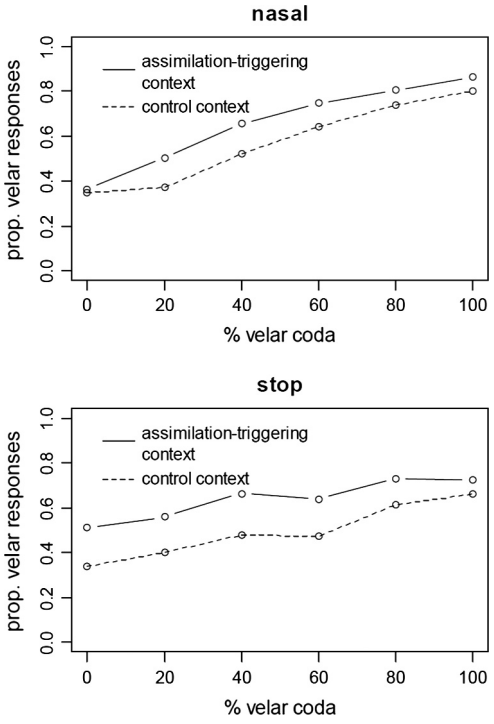

Fig. 9. Proportion of velar responses in Experiment 4 by English listeners. The top panel shows the results for nasals, the bottom panel the result for stops.

Listeners are likely to perceive an ambiguous sound as similar to an adjacent sound, if this adjacent sound is a good example of its category. Integrative and contrastive effects seem to compete in speech perception, and it is not yet clear under what circumstances which context effect prevails (Sjerps, Mitterer, & McQueen, 2012). Nevertheless, the results of Experiments 3 and 4 seem to suggest that the integrative effect would be the default effect when listeners with no language-specific knowledge hear a labial-to-velar continuum before a velar, and only with language experience, one shows a contrastive effect.

As was the case with Dutch listeners, English listeners showed no difference in slope between the assimilation-triggering and the control contexts ($b = -0.009$, $z = -0.5$, $p > 0.2$), but an overall effect of continuum ($b = 0.23$, $z = 18.28$, $p < 0.001$) when Continuum was used as a covariate rather than a factor. That is, the English listeners also fail to show compensation for assimilation in terms of both reduction of identification function slope and bias towards the velar end of the continuum.

General discussion

The current paper investigated how listeners would compensate for an optional labial-to-velar place assimilation rule in Korean that often leads to a categorical change in the speech output—i.e., the case in which the assimilated form does not carry any residual phonetic cues for its underlying phoneme. This form of assimilation has been

Table 5
Final model for the data of Experiment 4 (English participants).

Source	Regression weight	Estimate (SE)	*z*	*p*
Intercept	Context = Control, Manner = Nasal, Contimuum Step = 1	−0.61 (0.11)	−4.31	<0.001
Continuum	ContinuumStep = 2	0.36 (0.14)	2.66	<0.01
	ContinuumStep = 3	1 (0.14)	7.34	<0.001
	ContinuumStep = 4	1.49 (0.14)	10.56	<0.001
	ContinuumStep = 5	1.91 (0.15)	12.84	<0.001
	ContinuumStep = 6	2.3 (0.16)	14.51	<0.001
Context	Context = viable	0.54 (0.06)	9.34	<0.001
Manner	Manner = Stop	0.31 (0.14)	2.25	<0.05
Continuum × manner	ContinuumStep = 2:Manner = Stop	−0.13 (0.19)	−0.68	0.50
	ContinuumStep = 3:Manner = Stop	−0.38 (0.19)	−1.99	<0.05
	ContinuumStep = 4:Manner = Stop	−0.93 (0.19)	−4.79	<0.05
	ContinuumStep = 5:Manner = Stop	−0.83 (0.2)	−4.12	<0.01
	ContinuumStep = 6:Manner = Stop	−1.12 (0.21)	−5.30	<0.01

under investigation previously, with some puzzling results. Cho and McQueen (2008) presented two important findings. First, Korean listeners found it harder to recover the underlying phoneme in the labial-to-velar assimilation than in the alveolar-to-velar assimilation, showing an asymmetry in processing similar place assimilations. Second, despite the asymmetry, Korean listeners were still able to recover the underlying representation of the assimilated sound in both assimilatory environments, even though their speech materials in their study were created with no phonetic residuals of their underlying representations for both types of assimilation. (See also Cho & McQueen, 2011, for a case in which Korean listeners recovered a deleted phoneme due to consonant cluster simplification in Korean even when there was no bottom-up phonetic support.) These results are hard to be interpretable under the feature parsing account, as it does not predict such an asymmetrical effect when there are no residual phonetic cues. Nevertheless, Gow and Im (2004) implied that there might be no compensation process for this type of assimilation (with no bottom-up phonetic cues).

The first two experiments of the present study indicated that this is not the case. Korean listeners actively compensated for labial-to-velar place assimilation, making use of the context. That is, the alteration of place of articulation due to place assimilation was easier to overcome when the change occurred in the context of labial-to-velar assimilation that licensed it than in a neutral context. Experiment 1 showed this in a word-recognition task with an eye-tracking paradigm. In the critical trials, there was a mismatch between the target word to be clicked on (which ended with a labial) and the spoken word (which ended with a velar) in the instruction sentence stimuli that listeners actually heard. More looks and faster reactions to the labial target were observed when the mismatched 'spoken' velar was heard in the context that licensed the labial–velar assimilation than in the control context. This indicates that listeners indeed made use of the following context in recovering the underlying representation of the phonologically altered segment.

The results of Experiment 1 also have a methodological implication. The pattern of eye-movements was stable over the course of the experiments, while the context effects on RT were observed only in the first half of the experiment. Apparently, participants adapted to the mismatch between the targets and the spoken stimuli over time. That is, they encountered mismatched cases often enough as the experiment continued, so that they became less sensitive to the mismatch and faithful to the instruction. This strategic effect appears to have obliterated the phonological context effect observed in the first half of the experiment. This strategic adaptation, however, did not influence the eye-movements, indicating the effectiveness of measuring eye-movements in a word-recognition task in investigating various effects on spoken-word recognition processes which might otherwise be diluted by strategic, task-specific effects.

Experiment 2 revealed a similar result with a different method. This experiment employed a 2AFC phonetic categorization task, so that participants were never forced to overcome a mismatch between the target word and its corresponding spoken stimulus in Experiment 1. In this experiment, we also observed that velar–velar sequences were interpreted as possible intended labial–velar sequences in which the first consonant had undergone assimilation. Interestingly, however, this context effect was observed only when the stimuli contained strong velar tokens on the extreme end of the labial–velar continuum, which is quite unusual for context effects in speech perception. In the literature on context effects in speech perception (see, e.g., Beddor, Harnsberger, & Lindemann, 2002; Fowler, Brown, & Mann, 2000; Kingston, Kawahara, Chambless, Mash, & Brenner-Alsop, 2009; Lotto & Kluender, 1998; Mann, 1980; Mitterer, 2006a, 2006b; Smits, 2001), stronger context effects have been observed when the stimuli are more ambiguous, generally falling on the middle of the continuum. The opposite was observed here in the present study, with the context effect arising with the least ambiguous stimulus at the end of the labial-to-velar continuum.

This indicates that the current type of context effect may be different in nature from other context effects that have often been reported in the literature. Classical examples of context effects in speech perception, such as normalization for speaking rate (Newman & Sawusch, 1996) and for phonemic contexts (Mann, 1980) are often assumed to arise early in the perceptual process (Holt & Lotto, 2008) and are linked to uncertainty about the identity of the target phoneme as has been evident in robust context effects for

the most ambiguous targets. The different pattern for the current context effect may be taken as an indication that this arises at a later stage in the perceptual process, so that in fact *certainty* about the identity of the target phoneme is important for the context effect to arise.

Experiment 2 also was tailored to allow running a similar experiment with "naive" listeners with no knowledge of Korean. Experiment 3 consequently tested the perception of the same sequences by Dutch listeners. They showed a completely different pattern, with an integrative rather than a contrastive phonological effect triggered by the velar context—they categorized sounds more often as velars before a velar context than before a control context. Experiment 4 replicated this finding with English listeners, who are also unfamiliar with labial-to-velar assimilation, but are, in contrast to the Dutch listener group, familiar with unreleased stops. The similarity of these two groups, despite their different language backgrounds with regard to stop production, lend support to our claim that the contrastive context effect observed with Korean listeners is indeed due to their experience with labial-to-velar place assimilation in their native language, while an integrative effect arises with listeners who have no phonological experience with the labial-to-velar assimilation.

What does this mean for the different theoretical accounts for compensation for assimilation? The results clearly favor the phonological-inference account (Gaskell, 2003) over the other accounts such as the perceptual-integration account (Mitterer, Csépe, & Blomert, 2006) and the feature-parsing account (Gow, 2003). In line with the predictions of the phonological-inference account, Korean listeners made use of context information in compensation for labial-to-velar place assimilation, even when there were no bottom-up phonetic cues for the underlying target. In particular, the fact that the phonetic categorization by Korean listeners showed phonological context effects only when the stimuli were most clear velar–velar sequences appears to strengthen the phonological inference account. Given that labial-to-velar assimilation occurs most often in a categorical fashion, Korean listeners might have developed their phonological knowledge about the labial-to-velar assimilation that it frequently results in a completely assimilated form. Korean listeners are then likely to apply the phonological knowledge more efficiently to the 'complete' velar stimuli than ambiguous ones. This may explain why the contrastive phonological context effect in the phonetic categorization was maximal for the most clear velar stimuli. Finally, the fact that both Dutch and English listeners did not show contrastive context effects, but rather an integrative context effect in Experiment 3 and 4, again supports the phonological inference account—i.e., Dutch and English listeners did not have phonological knowledge of the labial–velar assimilation and therefore showed qualitatively different effects.

The results of the present study have therefore taken support away from the feature-parsing account, especially countering two important predictions—that is., the context effect is to be observed only when the assimilated segment contains phonetic residual cues for the underlying target, and no language-specific knowledge is required in processing assimilated forms regardless of whether the assimilation is gradient or complete, as particularly argued in Gow and Im (2004). The perceptual integration account is challenged by the results of the present study as well. While it predicts that the language-specific phonological knowledge can strengthen a context effect in line with the results of Experiment 1, it is disfavored by the findings in Experiments 2–4. First, the context effect only on clear velar–velar sequences in the phonetic categorization task in Experiment 2 runs counter to its prediction that the context effect occurs at the auditory level, which would be more likely reflected with ambiguous stimuli. Second, the perceptual integration account predicts that phonological context effects occur language-independently, while the language-dependent phonological knowledge just reinforces the context effect. But the qualitatively different context effects found for Korean versus Dutch/English listeners again runs counter to this prediction.

Does this mean that the phonological inference account is the most plausible account for compensation for assimilation? While it accounts most successfully for the kind of assimilation that has been tested in the present study (i.e., Korean labial-to-velar place assimilation), it may not be able to account for all forms of assimilation that have been investigated previously. Most of the previous investigations on compensation for assimilation have focused on assimilations which are likely to result in incompletely assimilated segments (presumably due to gestural overlaps), leaving fine-grained phonetic cues for the underlying phoneme in the signal. For these kinds of assimilation, all empirical data have shown that language-specific phonological knowledge does not play a primary role, but only modifies a language-independent perceptual bias. Such results were observed with nasal place assimilation (Mitterer, 2003), coronal place assimilation (Darcy et al., 2007), voicing assimilation in Hungarian (Gow & Im, 2004), and a Hungarian manner assimilation rule (Mitterer, Csépe, & Blomert, 2006; Mitterer, Csépe, Honbolygo, et al., 2006). These forms of assimilation can be construed as similar in nature to coarticulation in that both cases involve a phonetic change in a gradient fashion. It is therefore noteworthy that compensation for coarticulation has also been observed independently of language experience. One of the classic examples of compensation for coarticulation is that native listeners of American English perceive an ambiguous sound in the /da/-/ga/ continuum more likely as /ga/ after hearing /l/ than after hearing /r/, compensating for coarticulation that arise with /l/ versus /r/ (Mann, 1980). Although Japanese listeners have no experience with /r/-/l/ contrast in their native language (i.e., /r/ and /l/ are not phonologically contrastive), they show a comparable perceptual compensation for coarticulation like English listeners (Mann (1986). Perhaps more surprisingly, Lotto et al. (1997) found that even birds (i.e., Japanese quail) show evidence of compensation. This shows that context effects in speech perception can arise in the absence of language experience, if there are phonetic cues available in the input.

The phonological-inference account has no means to explain these language-independent effects that have been observed cross-linguistically across different assimilation and coarticulation types, and as such does not work for all types of assimilation. Phonological-inference, instead, seems play

a pivotal role in processing some, but not all, forms of assimilation, while mechanisms proposed by the feature parsing and the perceptual integration accounts for compensation may operate on other types of assimilation.

A question that follows is then why certain types of assimilation employ different compensatory mechanisms. In order to answer this question, we also need to consider yet another form of assimilation. It has been found in the literature that obligatory phonological rules which result in categorical outcomes do not generally challenge spoken word recognition. For example, German back fricatives are produced as either a velar [x] or as a palatal [ç], determined by the place of articulation of the preceding vowel (e.g., *Bach* produced as [bɑx] after a back vowel, but *Brecht* as [brɛçt] after a front vowel). Weber (2001) investigated how native and non-native listeners process this form of assimilation. She found that a violation of the assimilation rule leads to a "pop-out" for native listeners of German, but not for Dutch listeners, demonstrating the crucial role of the language-specific phonological knowledge in processing assimilation driven by a obligatory rule. Otake, Yoneyama, Cutler, and van der Lugt (1996) also found language-dependent effects for the obligatory assimilation of moraic nasals in Japanese. The Japanese moraic nasal has no fixed place of articulation, but obligatorily assimilates to the place of articulation of the following consonant. As the German fricative assimilation, this form of assimilation does not burden word recognition as the place of articulation of a nasal can always be determined by the place of articulation of the following consonant. Again, only Japanese listeners were able to make use of the place of articulation of the moraic nasal to predict the upcoming segment, while Dutch listeners could not.

To summarize, the results of the available studies taken together indicate that the perception of optional rules that often results in partial phonetic alteration (presumably due to gestural overlap) is prone to be language-independent, while obligatory rules resulting in categorical changes often lead to language-specific effects. The results of the present study, however, present a new case which requires use of language-specific knowledge in processing assimilation driven not by an obligatory rule but by an optional rule which nevertheless often leads to complete assimilation unlike many other optional assimilations.

Taking into account all the available data, we can now start to formulate *a priori* working hypothesis about how similarly or differently listeners compensate for different types of assimilation. The present study has clearly shown that the mechanisms of the phonological inference account most successfully explain the Korean listeners' behaviors of compensating for complete labial-to-velar assimilation, which has turned out to be language-specific, counter to underlying assumptions of the cross-linguistically applicable feature parsing theory or the perceptual integration account. It is nevertheless important to understand how language-independent compensation effects work differently from the language-specific effects. Language-independent effects are frequently observed across different types of assimilation cross-linguistically. These effects are different from the kind of assimilation observed in the present study in that they are either driven by fine-grained phonetic details that remain in the phonologically altered form (in line with the feature parsing account) or driven by listener-oriented perceptual motivations (in line with the perceptual integration account). However, the labial-to-velar place assimilation as found in Korean is typologically rare. This leads us to assume that it is not phonetically or perceptually motivated and the burden is completely on the part of the listener to recover the intended meaning. After all, to obtain a better and more comprehensive understanding of complicated listeners' behavior spoken word recognition, we need a model which integrates different mechanisms under different assimilation processes in a unified way.

Acknowledgments

We thank Jiseung Kim and Bobae Lee for the help with Korean and English data acquisition and Zhou Fang for Dutch data acquisition. We also thank Jiyoun Choi for helping to implement the experiment in the Netherlands and Helen Buckler for comments on an earlier version of this manuscript. This paper has greatly benefited from constructive suggestions and comments by Jim Magnuson and three anonymous reviewers. This work was supported by the National Research Foundation of Korea Grant funded by the Korean Government (NRF-2010-327-A00207).

Appendix 1

See Table A1.

Appendix 2

Results of the analyses of the eye-tracking data with fixation proportions.

In the present study, the eye-tracking data were analyzed using the Euclidean distance between fixation location and the center of the printed word as the dependent measure, rather than using fixation proportions which have been commonly used in eye-tracking studies with a visual-world paradigm. In this appendix, we report results of the fixation proportion analyses to show that the results from using two different methods are essentially the same.

As discussed in the main text, using the fixation proportions generates the issue to determine when a fixation should actually be counted as being *on* a given target, given that the current experimental design involves target-shape combinations as visual stimuli. This therefore should allow for a somewhat arbitrary choice for what counts as a fixation on the target. A loose, but sometimes employed, interpretation in the visual-world paradigm with four objects on the screen is that any fixation made in the same quadrant as an object is counted as a fixation on that object. However, a strict criterion would be that a fixation has to be actually on the objects itself (with only a small margin of error, i.e., 20 pixels). For the sake of completeness, results from using both the loose and the strict criteria are reported below.

The eye-tracking data for these trials are depicted in Fig. A1 for the loose criterion and Fig. A2 for the strict criterion. The figures show the fixations on the eventual target, its competitor (the other member of the minimal pair) and

Table A1
List of target word pairs with IPA symbols and English glosses.

Words with a stop coda (Velar /k/–Labial /p/)		Words with a nasal coda (Velar/ŋ/–Labial /m/)	
IPA	Gloss	IPA	Gloss
/kak/–/kap/	‘angle – pack’	/kjʌltʃʌŋ/–/kjʌltʃʌm/	‘decision – fault’
/kantʃʰʌk/–/kantʃʰʌp/	‘reclamation – spy’	/kjʌlhaŋ/–/kjʌlham/	‘flight cancellation – defect’
/kjʌk/–/kjʌp/	‘class – layer’	/koŋ/–/kom/	‘ball – bear’
/kuʃik/–/kuʃip/	‘old-style – ninety’	/kuɾiŋ/–/kuɾim/	‘hill – cloud’
/kuk/–/kup/	‘soup – heel’	/naŋ/–/nam/	‘pouch – south’
/kɨk/–/kɨp/	‘extreme – rank’	/noŋ/–/nom/	‘joke – guy’
/kiak/–/kiap/	‘instrumental music – pressure’	/taŋ/–/tam/	‘party – wall’
/kiʌk/–/kiʌp/	‘memory – corporation’	/mataŋ/–/matam/	‘garden – madame’
/pak/–/pap/	‘gourd – steamed rice’	/moŋ/–/mom/	‘dream – body’
/saŋkɨk/–/saŋkɨp/	‘extreme opposite – high rank’	/panɨŋ/–/panɨm/	‘reaction – semitone’
/suik/–/suip/	‘profit – import’	/paŋ/–/pam/	‘room – night’
/sutʃik/–/sutʃip/	‘verticality – collection’	/poŋ/–/pom/	‘pole – spring’
/sutʃʰʌk/–/sutʃʰʌp/	‘emaciation – notebook’	/puɨŋ/–/puɨm/	‘response – a notice of death’
/ʃik/–/ʃip/	‘ceremony – ten’	/saŋ/–/sam/	‘prize – three’
/ʌk/–/ʌp/	‘100 million – occupation’	/sɛŋ/–/sɛm/	‘life – spring’
/jʌk/–/jʌp/	‘station – leaf’	/sʌŋ/–/sʌm/	‘castle – island’
/tʌɨk/–/tʃɨp/	‘namely – juice’	/sujʌŋ/–/sujʌm/	‘swimming – moustache’
/tʃikak/–/tʃikap/	‘tardiness – purse’	/insaŋ/–/insam/	‘impression – ginseng’
/tʃikɨk/–/tʃikɨp/	‘utmost – payment’	/tʃaŋ/–/tʃam/	‘bowel – sleep’
/tʃijʌk/–/tʃijʌp/	‘area – minor details’	/tʃʌŋ/–/tʃʌm/	‘affection – point’
/tʃik/–/tʃip/	‘position – house’	/tʃoŋ/–/tʃom/	‘bell – moth’
/tʰok/–/tʰop/	‘pop-out – saw’	/tʃitʃʌŋ/–/tʃitʃʌm/	‘designation – branch’
/hak/–/hap/	‘crane – sum’	/tʃiŋ/–/tʃim/	‘gong – baggage’
/hok/–/hop/	‘bump – hop’	/tʃʰaŋ/–/tʃʰam/	‘window – truth’

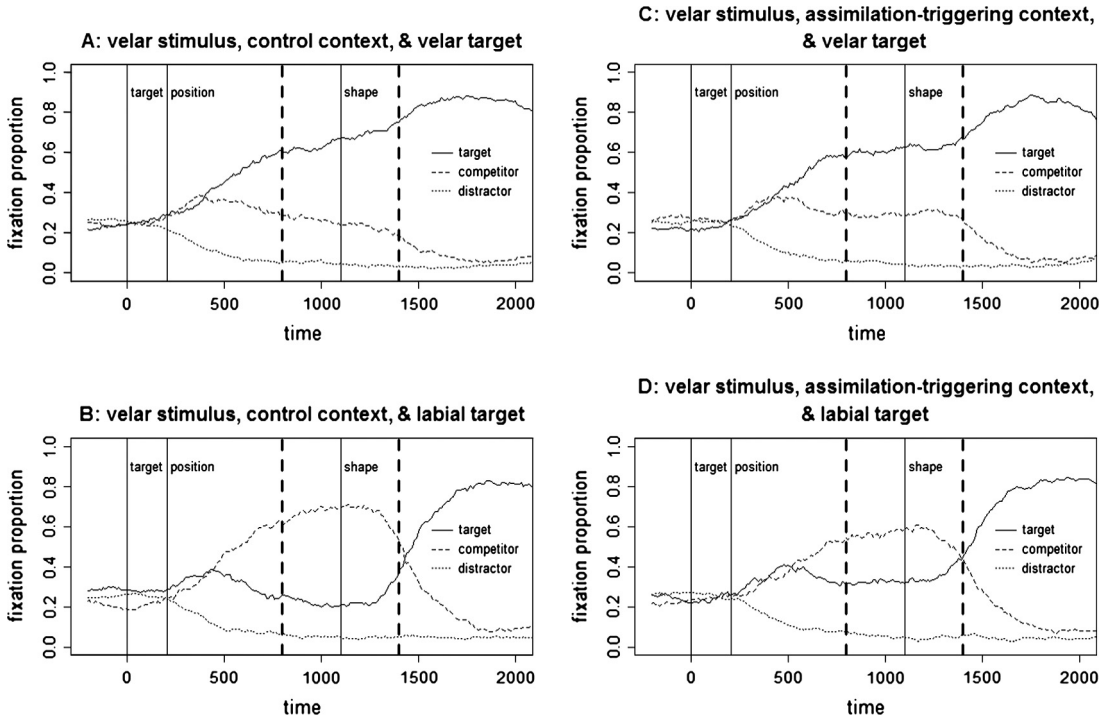

Fig. A1. Fixation proportions for the different screen objects (target, competitor, pooled distractors, coded with line thickness) for the experimental phonological trials. Fixations were based on quadrants, that is, any fixation in the upper right quadrant was counted as a fixation on the object in this quadrant. The thin lines indicate the onset of target, position, and object word in the sentence. The thick dashed lines indicate the critical time window (800–1400 ms) as determined from the semantic trials, on which the data analysis is based.

the two distractors. Note that the two distractors always differed in manner of articulation of the final consonant from the target. Hence, the distractors were easy to distinguish from the target. As the data show, participants were

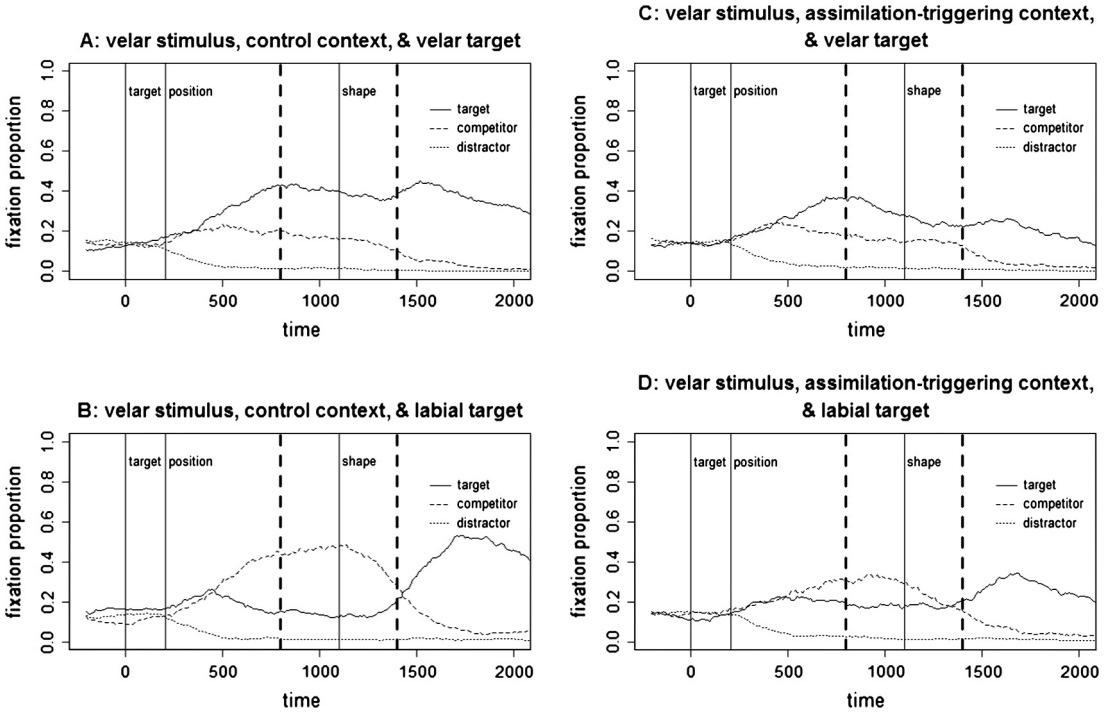

Fig. A2. Fixation proportions for the different screen objects (target, competitor, pooled distractors, coded with line thickness) for the experimental phonological trials. Fixations were considered to be on a word only if they were not further than 20 pixels away from the outer margin of the picture. The thin lines indicate the onset of target, position, and object word in the sentence. The thick dashed lines indicate the critical time window (800–1400 ms) as determined from the semantic trials, on which the data analysis is based.

mostly influenced by the surface place of articulation of the spoken word. That is, when they heard a velar-ending word, they looked at the corresponding printed word. This is the target on trials on which the shape information indicates to click on the velar-ending word (top panels of Figs. A1 and A2), but it is the competitor on trials on which the shape information indicates to click on the labial-ending word (lower panels of Figs. A1 and A2).

The overall effect was, however, moderated by the phonological context. In the assimilation-triggering context (the two right panels of Figs. A1 and A2), participants had less of a preference for the velar printed word than in the control context (the two left panels of Figs. A1 and A2). To statistically analyze this pattern of target preferences (that is, target fixations minus competitor fixations), we used a linear mixed effect model using the predictors Target Place and Context and their interaction. The target preference was calculated for the time window 800–1400 ms after the target onset, the time during which the context is already processed but the object information has not completely disambiguated the target yet, as estimated from the results obtained in the semantic trials. The fixation proportions were logistically transformed with ones and zero replaced by the mean of one and zero and the second highest or lowest possible fixation proportion respectively (Macmillan & Creelman, 1991).

The analysis revealed a main effect of Target Place (loose criterion: $b_{\text{TargetPlace = Velar}} = 6.81$, $p_{\text{MCMC}} = 0.0001$; strict criterion: $b_{\text{TargetPlace = Velar}} = 4.64$, $p_{\text{MCMC}} = 0.0001$), no overall effect of context (loose criterion: $b_{\text{Context = Assimilation-triggering}} = 0.5$, $p_{\text{MCMC}} = 0.06$; strict criterion: $b_{\text{Context = Assimilation-triggering}} = 0.35$, $p_{\text{MCMC}} = 0.08$) and an interaction of Target Place and Context (loose criterion: $b_{\text{Context = Assimilation-triggering} \times \text{TargetPlace = Velar}} = -2.5$, $p_{\text{MCMC}} = 0.0001$; strict criterion: $b_{\text{Context = Assimilation-triggering} \times \text{TargetPlace = Velar}} = -2.6$, $p_{\text{MCMC}} = 0.0001$). The negative beta weight indicates that the preference for the velar-bearing target (a positive effect) was less strong in the assimilation-triggering context. An analysis using fixation proportions hence reveals similar results as the one using Euclidean distances. This therefore suggests that the Euclidean distance measure may be used in replacement with the fixation proportion measure, especially in a case as the present study (with target-shape combinations as visual stimuli) in which using the fixation proportion measure would require an arbitrary criterion for what counts as a fixation on an object.

References

Aro, M., & Wimmer, H. (2003). Learning to read: English in comparison to six more regular orthographies. *Applied Psycholinguistics, 24*, 621–635. http://dx.doi.org/10.1017/S0142716403000316.

Beddor, P. S., Harnsberger, J. D., & Lindemann, S. (2002). Language-specific patterns of vowel-to-vowel coarticulation: Acoustic structures and their perceptual correlates. *Journal of Phonetics, 30*, 591–627. http://dx.doi.org/10.1006/jpho.2002.0177.

Blevins, J. (2004). *Evolutionary phonology: The emergence of sound patterns*. Cambridge: Cambridge University Press.

Boersma, P. (1998). *Functional phonology. Formalizing the interactions between articulatory and perceptual drives*. The Hague: Holland Academic Graphics.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10*, 433–436. http://dx.doi.org/10.1163/156856897X00357.

Cho, T., & McQueen, J. M. (2006). Phonological versus phonetic cues in native and non-native listening: Korean and Dutch listeners' perception of Dutch and English consonants. *Journal of the Acoustical Society of America, 119*, 3085–3096. http://dx.doi.org/10.1121/1.2188917.

Cho, T., & McQueen, J. M. (2008). Not all sounds in assimilation environment are perceived equally: Evidence from Korean. *Journal of Phonetics, 36*, 239–249. http://dx.doi.org/10.1016/j.wocn.2007.06.001.

Cho, T., & McQueen, J. M. (2011). Perceptual recovery from consonant-cluster simplification in Korean using language-specific phonological knowledge. *Journal of Psycholinguistic Research, 40*, 253–274.

Coenen, E., Zwitserlood, P., & Bölte, J. (2001). Consequences of assimilation for word recognition and lexical representation. *Language and Cognitive Processes, 16*, 535–564. http://dx.doi.org/10.1080/01690960143000155.

Darcy, I., Peperkamp, S., & Dupoux, E. (2007). Bilinguals play by the rules: Perceptual compensation for assimilation in late L2-learners. In J. Cole & J. I. Hualde (Eds.), *Laboratory phonology 9* (pp. 411–442). Berlin: Mouton de Gruyter.

Ernestus, M. (2000). *Voice assimilation and segment reduction in casual Dutch: A corpus-based study of the phonology–phonetics interface [dissertation]*. Utrecht, The Netherlands: LOT.

Fowler, C. A. (2006). Compensation for coarticulation reflects gesture perception, not spectral contrast. *Perception & Psychophysics, 68*, 161–177.

Fowler, C. A., Brown, J. M., & Mann, V. A. (2000). Contrast effects do not underlie effects of preceding liquids on stop-consonant identification by humans. *Journal of Experimental Psychology: Human Perception and Performance, 26*, 877–888. http://dx.doi.org/10.1037//0096-1523.26.3.877.

Gaskell, G., Clayards, M., & Niebuhr, O. (2009). *Cross-linguistic effects in the perception of assimilated speech*. Annual Meeting of the Psychonomics Society, Boston, MA.

Gaskell, M. G. (2003). Modeling regressive and progressive effects of assimilation in speech perception. *Journal of Phonetics, 31*, 447–463. http://dx.doi.org/10.1016/S0095-4470(03)00012-3.

Gaskell, M. G., & Marslen-Wilson, W. D. (1996). Phonological variation and inference in lexical access. *Journal of Experimental Psychology: Human Perception and Performance, 22*, 144–158. http://dx.doi.org/10.1037//0096-1523.22.1.144.

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.

Gaskell, M. G., & Marslen-Wilson, W. D. (2001). Lexical ambiguity resolution and spoken word recognition: Bridging the gap. *Journal of Memory and Language, 44*, 325–349. http://dx.doi.org/10.1006/jmla.2000.2741.

Gaskell, M. G., & Snoeren, N. D. (2008). The impact of strong assimilation on the perception of connected speech. *Journal of Experimental Psychology: Human Perception and Performance, 34*, 1632–1647. http://dx.doi.org/10.1037/a0011977.

Gow, D. W. (2002). Does English coronal place assimilation create lexical ambiguity. *Journal of Experimental Psychology: Human Perception and Performance, 28*, 163–179. http://dx.doi.org/10.1037//0096-1523.28.1.163.

Gow, D. W. (2003). Feature parsing: Feature cue mapping in spoken word recognition. *Perception & Psychophysics, 65*, 575–590. http://dx.doi.org/10.3758/BF03194584.

Gow, D. W., & Im, A. M. (2004). A cross-linguistic examination of assimilation context effects. *Journal of Memory and Language, 51*, 279–296. http://dx.doi.org/10.1016/j.jml.2004.05.004.

Holt, L. L., & Lotto, A. J. (2008). Speech perception within an auditory cognitive science framework. *Current Directions in Psychological Science, 17*, 42–46. http://dx.doi.org/10.1111/j.1467-8721.2008.00545.x.

Hura, S. L., Lindblom, B., & Diehl, R. (1992). On the role of perception in shaping phonological assimilation rules. *Language and Speech, 35*, 59–72.

Jun, J. (1996). Place Assimilation is not the result of Gestural Overlap: Evidence from Korean and English. *Phonology, 13*, 377–407.

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

Kawahara, H., Masuda-Katsuse, I., & de Cheveigné, A. (1999). Restructuring speech representations using a pitch-adaptive time-frequency smoothing and an instantaneous-frequency based F0 extraction. *Speech Communication, 27*, 187–207. http://dx.doi.org/10.1016/S0167-6393(98)00085-5.

Kim-Renaud, Young-Key (1975). *Korean consonantal phonology*. Seoul: Tower Press.

Kingston, J., Kawahara, S., Chambless, D., Mash, D., & Brenner-Alsop, E. (2009). Contextual effects on the perception of duration. *Journal of Phonetics, 37*, 297–320. http://dx.doi.org/10.1016/j.wocn.2009.03.007.

Kleiner, M., Brainard, D., & Pelli, D. (2007). What's new in Psychtoolbox-3? *Perception 36*, ECVP Abstract Supplement.

Kühnert, B., & Nolan, F. (1999). The origin of coarticulation. In W. J. Hardcastle & N. Hewlett (Eds.), *Coarticulation: Theory, data and techniques in speech production* (pp. 7–30). Cambridge, UK: Cambridge University Press.

Kuzla, C., Ernestus, M., & Mitterer, H. (2010). Compensation for assimilatory devoicing and prosodic structure in German fricative perception. In C. Fougeron, B. Kühnert, M. D'Imperio, & N. Vallée (Eds.), *Laboratory phonology 10* (pp. 731–758). Berlin: Mouton.

Lee, I., & Ramsey, S. R. (2000). *The Korean language*. Albany, NY: SUNY Press.

Lotto, A. J., & Kluender, K. R. (1998). General contrast effects in speech perception: Effect of preceding liquid on stop consonant identification. *Perception & Psychophysics, 60*, 602–619.

Lotto, A. J., Kluender, K. R., & Holt, L. L. (1997). Perceptual compensation for coarticulation by Japanese quail (*Coturnix coturnix japonica*). *Journal of the Acoustical Society of America, 102*, 1134–1140. http://dx.doi.org/10.1121/1.419865.

Macmillan, N. A., & Creelman, D. (1991). *Detection theory: A user's guide*. Oxford: Blackwell Publishers.

Maes, A., Arts, A., & Noordman, L. (2004). Reference management in instructive discourse. *Discourse Processes, 37*, 117–144. http://dx.doi.org/10.1207/s15326950dp3702_3.

Mann, V. A. (1980). Influence of preceding liquid on stop-consonant perception. *Perception & Psychophysics, 28*, 407–412.

Mann, V. A. (1986). Distinguishing universal and language-dependent levels of speech perception: Evidence from Japanese listeners perception of English 'l' and 'r'. *Cognition, 24*, 169–196. http://dx.doi.org/10.1016/S0010-0277(86)80001-4.

McMurray, B., Clayards, M. A., Tanenhaus, M. K., & Aslin, R. N. (2008). Tracking the time course of phonetic cue integration during spoken word recognition. *Psychonomic Bulletin & Review, 15*, 1064–1071. http://dx.doi.org/10.3758/pbr.15.6.1064.

McQueen, J. M., & Viebahn, M. (2007). Tracking recognition of spoken words by tracking looks to printed words. *Quarterly Journal of Experimental Psychology, 60*, 661–671. http://dx.doi.org/10.1121/1.419865.

Mitterer, H. (2003). *Understanding "gardem bench": Studies on the perception of assimilation word forms [dissertation]*. Maastricht, The Netherlands: Universiteit Maastricht.

Mitterer, H. (2006a). Is vowel normalization independent of lexical processing? *Phonetica, 63*, 209–229. http://dx.doi.org/10.1159/000097306.

Mitterer, H. (2006b). On the causes of compensation for coarticulation: Evidence for phonological mediation. *Perception & Psychophysics, 68*, 1227–1240. http://dx.doi.org/10.3758/BF03193723.

Mitterer, H. (2011a). The mental lexicon is fully specified: Evidence from eye-tracking. *Journal of Experimental Psychology: Human Perception and Performance, 37*, 496–513. http://dx.doi.org/10.1037/a0020989.

Mitterer, H. (2011b). Recognizing reduced forms: Different processing mechanisms for similar reductions. *Journal of Phonetics, 39*, 298–303. http://dx.doi.org/10.1016/j.wocn.2010.11.009.

Mitterer, H., & Blomert, L. (2003). Coping with phonological assimilation in speech perception: Evidence for early compensation. *Perception & Psychophysics, 65*, 956–969. http://dx.doi.org/10.3758/BF03194826.

Mitterer, H., Chen, Y., & Zhou, X. (2011). Phonological abstraction in processing lexical-tone variation: Evidence from a learning paradigm. *Cognitive Science, 35*, 184–197. http://dx.doi.org/10.1111/j.1551-6709.2010.01140.x.

Mitterer, H., Csépe, V., & Blomert, L. (2006). The role of perceptual integration in the recognition of assimilated word forms. *Quarterly Journal of Experimental Psychology, 59*, 1395–1424. http://dx.doi.org/10.1080/17470210500198726.

Mitterer, H., Csépe, V., Honbolygo, F., & Blomert, L. (2006). The recognition of phonologically assimilated words does not depend on specific language experience. *Cognitive Science, 30*, 451–479. http://dx.doi.org/10.1207/s15516709cog0000_57.

Mitterer, H., & Ernestus, M. (2006). Listeners recover /t/s that speakers lenite: Evidence from /t/-lenition in Dutch. *Journal of Phonetics, 34*, 73–103. http://dx.doi.org/10.1016/j.wocn.2005.03.003.

Mitterer, H., & McQueen, J. M. (2009). Processing reduced word-forms in speech perception using probabilistic knowledge about speech production. *Journal of Experimental Psychology: Human Perception and Performance, 35*, 244–263. http://dx.doi.org/10.1037/a0012730.

Newman, R. S., & Sawusch, J. R. (1996). Perceptual normalization for speaking rate: Effects of temporal distance. *Perception & Psychophysics, 58*, 540–560. http://dx.doi.org/10.3758/BF03213089.

Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology, 47*, 204–238. http://dx.doi.org/10.1016/S0010-0285(03)00006-9.

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. http://dx.doi.org/10.1121/1.417239.

Poellmann, K., McQueen, J. M., & Mitterer, H. (2011). The time course of perceptual learning. In W. -S. Lee & E. Zee (Eds.), Proceedings of the 17th international congress of phonetic sciences 2011 [ICPhS XVII] (pp. 1618–1621). Hong Kong: Department of Chinese, Translation and Linguistics, City University of Hong Kong.

Salverda, A. P., & Tanenhaus, M. K. (2010). Tracking the time course of orthographic information in spoken-word recognition. *Journal of Experimental Psychology: Learning Memory and Cognition, 36*, 1108–1117. http://dx.doi.org/10.1037/a0019901.

Sjerps, M., Mitterer, H., & McQueen, J. M. (2012). Hemispheric differences in the effects of context on vowel perception. *Brain and Language, 120*, 401–405. http://dx.doi.org/10.1016/j.bandl.2011.12.012.

Smits, R. (2001). Evidence for hierarchical categorization of coarticulated phonemes. *Journal of Experimental Psychology: Human Perception and Performance, 27*, 1145–1162. http://dx.doi.org/10.1037//0096-1523.27.5.1145.

Snoeren, N. D., Segui, J., & Hallé, P. A. (2008). Perceptual processing of partially and fully assimilated words in French. *Journal of Experimental Psychology: Human Perception and Performance, 34*, 193–204. http://dx.doi.org/10.1037/0096-1523.34.1.193.

Son, M., Kochetov, A., & Pouplier, M. (2007). The role of gestural overlap in perceptual place assimilation in Korean. In J. Cole & J. I. Hualde (Eds.). *Papers in laboratory phonology* (Vol. 9, pp. 504–534). Berlin: Mouton de Gruyter.

Steriade, D. (2001). Directional asymmetries in place assimilation: A perceptual account. In E. Hume & K. Johnson (Eds.), *The role of speech perception in phonology* (pp. 219–250). New York, NJ: Academic Press.

Tavabi, K., Elling, L., Dobel, C., Pantev, C., & Zwitserlood, P. (2009). Effects of place of articulation changes on auditory neural activity: A magnetoencephalography study. *Plos One, 4*. http://dx.doi.org/10.1371/journal.pone.0004452.

Turennout, M. v., Hagoort, P., & Brown, C. M. (1998). Brain activity during speaking: From syntax to phonology in 40 milliseconds. *Science, 280*, 572–574. http://dx.doi.org/10.1126/science.280.5363.572.

Van der Heijden, A. H. C., Hagenaar, R., & Bloem, W. (1984). Two stages in postcategorial filtering and selection. *Memory and Cognition, 12*, 458–469. http://dx.doi.org/10.3758/BF03198307.

Viswanathan, N., Magnuson, J. S., & Fowler, C. A. (2010). Compensation for coarticulation: Disentangling auditory and gestural theories of perception of coarticulatory effects in speech. *Journal of Experimental Psychology: Human Perception and Performance, 36*, 1005–1015. http://dx.doi.org/10.1037/a0018391.

Weber, A. (2001). Help or hindrance: How violation of different assimilation rules affects spoken-language processing. *Language and Speech, 44*, 95–118. http://dx.doi.org/10.1177/00238309010440010401.