Perceptual Recovery from Consonant-Cluster Simplification in Korean Using Language-Specific Phonological Knowledge

Taehong Cho · James M. McQueen

Abstract Two experiments examined whether perceptual recovery from Korean consonant-cluster simplification is based on language-specific phonological knowledge. In tri-consonantal C1C2C3 sequences such as /lkt/ and /lpt/ in Seoul Korean, either C1 or C2 can be completely deleted. Seoul Koreans monitored for C2 targets (/p/ or / k/, deleted or preserved) in the second word of a two-word phrase with an underlying /l/-C2-/t/ sequence. In Experiment 1 the target-bearing words had contextual lexical-semantic support. Listeners recovered deleted targets as fast and as accurately as preserved targets with both Word and Intonational Phrase (IP) boundaries between the two words. In Experiment 2, contexts were low-pass filtered. Listeners were still able to recover deleted targets as well as preserved targets in IP-boundary contexts, but better with physically-present targets than with deleted targets in Word-boundary contexts. This suggests that the benefit of having target acoustic-phonetic information emerges only when higher-order (contextual and phrase-boundary) information is not available. The strikingly efficient recovery of deleted phonemes with neither acoustic-phonetic cues nor contextual support demonstrates that language-specific phonological knowledge, rather than language-universal perceptual processes which rely on fine-grained phonetic details, is employed when the listener perceives the results of a continuous-speech process in which reduction is phonetically complete.

Keywords Speech perception · Continuous-speech processes · Probabilistic phonological knowledge · Consonant-cluster simplification · Korean · Phonetic information · Prosody

T. Cho (✉)
Department of English Language and Literature, Hanyang University, 17 Haengdang-dong,
Seongdong-gu, Seoul 133-791, Korea
e-mail: tcho@hanyang.ac.kr

J. M. McQueen
Behavioural Science Institute and Donders Institute for Brain, Cognition and Behaviour,
Centre for Cognition, Radboud University Nijmegen, Nijmegen, The Netherlands

J. M. McQueen
Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands

Introduction

To recognize spoken words, listeners must map acoustic-phonetic patterns onto phonological representations in the mental lexicon. This mapping is already very complex when the surface patterns correspond faithfully to the underlying representations (see McQueen 2007 for review). Frequently, however, this mapping becomes even more complex—in cases where there is no one-to-one correspondence between surface realizations and words' citation forms. Several processes which operate during the production of continuous speech, especially when it is fast and casual, can alter how sounds in words are realized. They include reduction (Kohler 2000; Mitterer and Ernestus 2006; Mitterer and McQueen 2009), epenthesis (van Donselaar et al. 1999), and liaison (Spinelli et al. 2003). Many of them are variable (i.e., not mandatory): They occur more often in some speech styles or words. How do listeners recover speakers' intended words when the sounds in those words have been mangled by continuous-speech processes?

One answer to this question is that listeners have knowledge about these processes, and thus, during recognition, can recover from adjustments that speakers made during production. A widely-studied process has been assimilation. In place assimilation in English, for example, words ending with coronal consonants (e.g., the /t/ in *late*) can sound like they end with a velar consonant when followed by a word beginning with a velar (e.g., *late cruise* sounding like *lake cruise*). Listeners can nonetheless recognize an assimilated word, but only if it appears with a following context which licenses the assimilation (Coenen et al. 2001; Gaskell and Marslen-Wilson 1996, 1998; Gow 2001, 2002). This suggests that phonological knowledge about place assimilation can be used to recover from its effects. Gaskell and Marslen-Wilson (1996), for example, proposed a process of phonological inference.

Another answer, however, is that listeners can find remnants of the speaker's intentions in the acoustics; that is, the speech-production mangle does not completely destroy the components of the underlying message. Place assimilation in English, for example, can be phonetically incomplete, such that an assimilated consonant may maintain some of its identity (e.g., the final consonant in *late*, when assimilated before *cruise*, has some velar and some coronal features). Listeners are sensitive to these fine-grained differences, and can use them in lexical disambiguation (e.g., distinguishing between *late* and *lake*; Gow 2002).

Given these acoustic signatures, it is possible that listeners could recover from continuous-speech phenomena without recourse to phonological knowledge. Participants listening to a language they are unfamiliar with can indeed show the same perceptual sensitivity to the results of assimilation as native speakers of that language (Gow and Im 2004; Mitterer et al. 2006). Language-universal perceptual processes can thus operate on the acoustic evidence to retrieve underlying sounds. Differences in perception between native and nonnative listeners in other studies nevertheless suggest that language-specific knowledge is used to deal with assimilation (Otake et al. 1996; Weber 2001). An important issue, therefore, is to establish when phonological knowledge is or is not used to recover from the effects of continuous-speech processes.

When a tri-consonantal sequence C1C2C3 in Korean occurs in a word-form due to morpheme concatenation, C1 or C2 can be deleted, especially when C1 is /l/ as in /lkt/ and /lpt/ clusters (Cho 1999; Kim-Renaud 1976). The entire segment can be eliminated during consonant-cluster simplification (Cho 1999; Cho and Kim 2009), leaving no residual phonetic cues to the underlying segment. A corpus study by Cho and Kim (2009) reported that speakers of Seoul Korean produced all three variants (with C1 or C2 deleted or both preserved) of /lkt/ and /lpt/ clusters, among which C2-deleted variants with no

residual phonetic information occurred more than 50% of the time and C1-deleted (therefore C2-preserved) variants 17.8% of the time. Cho and Kim suggested that both C2-deleted and C2-preserved simplified forms are phonologically permissible in Seoul Korean, showing intra-dialectal variation of cluster simplification. Successful recovery of a completely missing segment in a simplification environment is therefore likely to rely on knowledge about the phonological process (i.e., either C1 or C2 may be deleted), without reference to residual phonetic information. We tested here whether Korean listeners can indeed recover from consonant-cluster simplification especially when there are no fine phonetic remnants of the underlying phonemes, and hence the hypothesis that listeners deal with the effects of continuous-speech processes, at least in part, by using language-specific phonological knowledge.

Perceptual recovery from Korean consonant-cluster simplification, however, may not depend solely on language-specific phonological knowledge. Listeners may also use contextual lexical-semantic information to infer what the simplified word is (see e.g., Dahan and Tanenhaus 2004; Mattys et al. 2005; Tyler and Wessels 1983). Once the intended word has been inferred, recovery of the deleted phoneme could proceed without recourse to phonological knowledge about the consonant-cluster simplification process itself. We therefore tested the phonological knowledge hypothesis in two phoneme-monitoring experiments in which the critical stimuli were presented either with or without contextual lexical-semantic support. In both experiments we presented Seoul Korean listeners with two-word utterances. The first word either was an acoustically clear token (Experiment 1) or was acoustically filtered (Experiment 2). The second word contained an underlying tri-consonantal sequence and the target (either /p/ or /k/) was C2. The sequence was presented with C2 deleted (C1 and C3 were still present) or with C2 preserved (C1 was deleted, but C3 was preserved).

In Experiment 1, we examined how listeners would process simplified forms of consonant cluster simplification with contextually-driven lexical-semantic information available in the stimuli, as in a phrase not^hi-lil $il\underline{\textbf{k}}$-$t\partial lato$ ('although reading the notebook'). The preceding contextual word (not^hi-lil, 'notebook-Accusative') was semantically related to the target-bearing word ($il\underline{\textbf{k}}$-$t\partial lato$, 'to read-conjunction'). Given that listeners generally process simplified forms in context, we first tested how efficiently listeners would restore deleted phonemes as compared to physically-present phonemes in such an environment. In Experiment 2, the preceding contextual information was removed by low-pass filtering, so that we could test whether listeners would still be able to retrieve underlying representations of deleted phonemes efficiently even when both higher-order contextual support and lower-order acoustic-phonetic segmental cues are absent from the input.

Across the two experiments, we also tested a more specific hypothesis about the role of phonetic versus higher-order information in the perception of preserved versus deleted phoneme targets: Is perception of a deleted phoneme just as good as perception of a preserved phoneme? On the one hand, given that physically-present (preserved) targets carry direct acoustic-phonetic cues to consonant identity, listeners might be expected to detect physically-present phonemes more efficiently than physically-absent (deleted) phonemes. On the other hand, although deleted phonemes lack acoustic-phonetic segmental cues, higher-order information such as contextual and phonological information could be enough to match the benefit of the acoustic-phonetic cues in physically-present targets. Mattys et al. (2005) indeed suggested that, in lexical segmentation, when different levels of information are available, higher-order information (e.g., lexicality and semantic and syntactic context) is weighted more than sublexical information (e.g., segmental information). If this is the case in the

processing of simplified consonant clusters in Korean, listeners should detect physically-absent phonemes as efficiently as physically-present phonemes, especially in Experiment 1 when both contextual and phonological information are available. However, if phonological information alone is sufficient for listeners to be able to recover from simplification, they should detect phonemes as easily when they are deleted as when they are preserved even without contextual support, as in Experiment 2.

We also tested whether phoneme monitoring efficiency is conditioned by the size of the prosodic boundary before the target-bearing word. Christophe et al. (2004), in word and phoneme monitoring tasks in French, found that lexical access was faster when words (e.g., *chat*, cat) appeared in two-word sequences (e.g., *chat grincheux,* lit. cat grumpy) which spanned Phonological Phrase boundaries than in sequences within a single Phonological Phrase. Cho et al. (2007) found that word recognition was facilitated when phonetic cues to an Intonational Phrase (IP) boundary strengthened the evidence about the location of a critical word boundary. In a study of artificial language learning by Korean listeners, Kim (2003, 2004) showed that lexical segmentation was facilitated when acoustic cues signaled prosodic phrase boundaries (Accentual-Phrase boundaries; these are intermediate-level prosodic boundaries in Korean; Jun 1993, 2000). More recently, in a word-spotting experiment in Korean, Kim and Cho (2009) showed that Korean listeners benefited from the presence of a larger prosodic boundary (i.e., an IP boundary) in lexical segmentation. Listeners may thus be able to detect targets more readily when target-bearing words appear at a larger prosodic boundary. The materials were therefore realized with either an IP or a phrase-internal Word boundary between the words.

Finally, the present study tested whether listeners treat /k/ and /p/ differently. It has been suggested in the phonological literature that some sounds are perceptually more robust than others, due to intrinsically different acoustic-phonetic properties, and that such differences may constrain certain aspects of the sound patterns of the world languages (Flemming 1995; Hayes et al. 2004; Hume and Johnson 2001; Jun 1995, 2004; Steriade 1999, 2001). For example, according to the listener-oriented production hypothesis (Jun 1995, 2004; and see references therein), speakers make more effort to preserve sounds with more robust acoustic cues, and they do this for the listener's benefit. This speech-production effect has been suggested as an explanation for the cross-linguistic tendency that velars undergo phonological alterations less often than labials or alveolars: Velars are perceptual more salient (presumably with compactness and convergence of F2 and F3 as well as a comparatively long formant transition due to sluggishness of the tongue body movement) than labials or alveolars, such that speakers preserve /k/ for the listener's benefit, and listeners in turn take advantage of that.

Cho and McQueen (2008) created phonologically illegal assimilatory cases in Korean in which /k/ and /p/ were illegally assimilated to the following /t/, and tested whether listeners would react differently to assimilated /k/ versus /p/, relative to the unassimilated /k/ and /p/. Their results showed that detection speed and accuracy were higher for the physically-present /k/ than for the physically-present /p/, relative to their assimilated counterparts, suggesting that listeners indeed benefited more from the acoustic-phonetic properties of /k/ than from those of /p/. The stimuli in the present study also contained /k/ and /p/ which were either physically present or absent. If /k/ carries more robust acoustic phonetic cues than /p/ does, and listeners benefit more from the presence of a /k/ than from the presence of a /p/, detection of physically-present /k/ (relative to deleted /k/) should be easier than detection of physically-present /p/ (relative to deleted /p/).

Experiment 1

In this experiment, we examined how listeners recover underlying representations of physically-absent (deleted) phonemes due to consonant cluster simplification with contextually-driven lexical-semantic information available in the stimuli.

Method

Participants

Ninety-six students at Hanyang University, Seoul, were paid to participate. They were all native speakers of Korean, who had all been taught English as a second language during secondary education.

Materials

The target phoneme in experimental trials was always the second member of the complex coda (C1C2) in the first syllable of the second word of a two-word utterance (e.g., for /p/, *sija-ka nəl**p**-təlato* 'although the view is wide'; for /k/, *koŋki-ka mal**k**-təlato*, 'although the air is clear'; note that '-' refers to a morpheme boundary, and that the words are specified phonemically with IPA transcriptions). In these trials C1 was always /l/, and C3 (the initial consonant of the second morpheme) was always /t/. Six words served as /p/-bearing and eight as /k/-bearing test words, yielding a total of 14 test words. The full set, which consists of all lexical sequences in Korean meeting the required constraints that we could find, is in the Appendix.

Another 14 two-word sequences were constructed as foils which had no instances of the specified target (thus no response was required). The second words in the foil sequences were the same as the test words but the first (preceding) words were different from those used in the experimental items. The presence of these foils prevented participants from predicting reliably whether a given trial would contain a target. The foil trials always appeared later in the experiment than their corresponding experimental trials with the same word, so that they would not influence the listener's performance on the experimental items. Furthermore, 20 fillers and 20 filler foils (again two-word phrases) were constructed with tri-consonantal sequences not used in the test words.

In total, there were 68 trials with /p/ and /k/ as targets (34 target-present and 34 foil trials). All targets occurred in similar contexts in terms of prosodic boundaries and target locations. Finally, 15 practice items were constructed. The practice items included both C2-deleted and C2-preserved items with consonant clusters not used in the test words.

Materials were recorded in a sound-damped booth by a male native Korean speaker. IP renditions were obtained by instructing the speaker to produce the stimuli naturally and comfortably. For Word boundary renditions, the speaker was asked to connect the two words as if forming a single unit. Each test sentence was read several times, with either an IP boundary or a phrase-internal Word boundary between the two words. The prosodic boundaries (IP vs. Word) of the spoken stimuli were transcribed by two trained Korean phoneticians (one of them was the first author). Among those tokens that were agreed by the transcribers to meet the intended prosodic boundary conditions, the best token of each stimulus (the one that was perceived to be most naturally pronounced as intended) was chosen for use in the experiments.

Note that, in Korean, C3 (/t/) is always released into the following vowel, while C2 (/p/ or /k/) is never released. Therefore, it was either entirely absent (in the C2-deleted condition), or unreleased (in the C2-preserved condition), as appropriate for coda stops in Korean (Cho and McQueen 2006). C2 targets, when preserved, were therefore cued primarily by acoustic information, especially formant transitions, in the preceding vowels. Given that it was critical for the present study to obtain completely deleted C2 forms, three steps were taken.

First, the speaker read two scripts, one with the target-bearing words written with C2 deleted, and one with C2 preserved (C1-deleted). That is, the target-bearing words were intentionally misspelled as either C2-deleted or C1-deleted (C2-preserved) forms in order to induce the speaker to pronounce completely deleted forms without being influenced by the written underlying consonant (C1 or C2) that was not to be pronounced. Since the Korean alphabet (*Hangul*) is phonologically fully transparent, the use of such scripts ensured that the speaker pronounced the test words as intended. Second, two trained Korean phoneticians examined the recorded C2-deleted variants, and reported that the selected C2-deleted variants did not contain any perceptual phonetic remnants of the underlying C2. Third, we checked that the C2s in the C2-deleted variants were completely deleted by running an auditory-perceptual pretest with a two-alternative forced choice (2AFC) task.

Pretest

Materials

Test items contained only a part of each C2-deleted VC1C3V sequence. That part consisted of only the second half of the preceding vowel, then the C1C3 sequence, and then the first half of the following vowel.[1] The excised parts therefore included the critical portions, especially the preceding vowel and the liquid /l/ (C1), which would contain acoustic-phonetic information for the deleted C2, if it were to exist, but without lexical/contextual support. (Recall that, in Korean, syllable-final stops are never released, and therefore their acoustic manifestation is primarily reflected in formant transitions in the surrounding vowels; e.g., Cho and McQueen 2006). The auditory stimuli were excised from all 28 test items (6 /p/-deleted and 8 /k/-deleted items, each with IP and Word boundaries).

Procedure

Thirty Seoul Korean undergraduate students (other than those who participated in the main experiment) participated in the pretest. They were asked to choose whichever of the two consonants, /p/ or /k/, they thought could possibly be present between the two consonants C1 and C3 in each VC1C3V sequence. The stimuli were presented through headphones using NESU experimental control software. On each trial, the subject heard the critical part twice (with a 500 ms inter-stimulus interval) and had to press a response button labeled either /p/ ('ㅂ') or /k/ ('ㄱ'). So that the total number of /p/ targets and /k/ targets would be the same, the trials were repeated three times for the /p/-deleted variants and four times for the /k/-deleted variants, yielding 96 trials. The inter-trial interval was 4 s.

---

[1] In several cases it was difficult to excise exactly the second half of V1 because the boundary between the vowel and the following sonorant /l/ (C1) was hard to distinguish. Given that the tokens which were relatively easy to segment into V1 and C1 showed approximately a two-to-one ratio between V1 and C1, we decided in the more difficult cases to excise the part including the second two-thirds of the entire V1C1 sonorant sequence.

Results

A series of chance-level one-sample t-tests were conducted separately for /p/ and /k/ targets in each of the IP and Word conditions. Listeners performed the task at chance level, as revealed in both by-subjects and by-items analyses (all differences from 0.5 at p > 0.1). The mean percent correct accuracies were 50.9% (s.e., 1.4) and 50.3% (s.e., 1.7) for /k/ and /p/ targets in the IP condition and 49.0% (s.e., 2.0) and 51.1% (s.e., 2.1) for /k/ and /p/ targets in the Word condition. This suggests that the deletion of C2 was complete. The materials did not contain any acoustic residuals of the C2-deleted phonemes that could be used by listeners at the auditory level. Repeated measures ANOVAs with Boundary (IP vs. Word) and Consonant (/k/ vs. /p/) as between-subject factors also showed no significant main effects or interactions (Boundary: $F1(1, 29) < 1$, $F2(1, 12) < 1$; Consonant $F1(1, 29) < 1$, $F2(1, 12) < 1$). This confirmed that the auditory materials used in the present study were neutral, with no acoustic-perceptual bias toward any particular target in either condition.

Procedure

Participants in the main experiment were divided into four groups of 24. Two groups were presented with the IP-boundary and two with the Word-boundary materials. Two lists were constructed for each prosodic boundary condition, with the C2 deletion condition (C2-deleted or C2-preserved) counterbalanced over the two lists. One group of listeners within each prosodic boundary condition heard each list. Each listener heard all fillers and one version of each of the test words, but never both the C2-deleted and C2-preserved form of the same test item.

The participants' task was to detect pre-specified target phonemes, which they were told would occur, if at all, within the first syllable of the second word in each utterance. Targets were printed in *Hangul* for 1 s on a computer screen: ㅂ (p) and ㄱ (k). Each spoken stimulus began immediately after the visual target disappeared. A new target was presented prior to each stimulus. Participants were instructed to press a button as fast and as accurately as possible when they detected a target. Response latencies relative to the offset of the C1C2 sequence (i.e., the end of the stop closure for C3 /t/) and errors were recorded. (Since targets were not physically present in the C2-deleted condition, C3 was used as a consistent alignment point for measuring latency).

In order to ensure that the speaker had reliably produced utterances with Word and IP boundaries, we measured the duration of the vowel of the preboundary syllable, as phrase-final lengthening is one of the most reliable phonetic signatures of a prosodic boundary (e.g., for English, Edwards et al. 1991; Wightman et al. 1992; for Korean, Jun 1993; Cho and Keating 2001; Kim and Cho 2009). An ANOVA showed that there was a main effect of prosodic boundary on preboundary vowel duration, showing that preboundary vowel was significantly longer before IP than Word boundary ($F(1, 12) = 778.64$, p < 0.0001; means, 221.6 ms vs. 87.5 ms). There were no effects of C2 deletion, C2 identity or interactions between factors on preboundary vowel duration. These results thus provided acoustic support for the IP versus Word boundary divisions that were determined by the two transcribers.

Results

Each subject was expected to provide at least two correct responses for either /p/ or /k/ in the C2-preserved condition. One subject failed this criterion. Table 1 shows mean error rates and latencies for the remaining 95 subjects.

Table 1 Mean proportion of errors (%) and mean Reaction Times (RT) in Experiment 1 (with context), for phoneme monitoring responses to the targets (C2) /k/ and /p/ in underlyingly triconsonantal (C1C2C3) sequences

Measures	Prosodic boundary	C2 = /k/		C2 = /p/		Overall (/k/ & /p/ combined)		
		Deleted	Preserved	Deleted	Preserved	Deleted	Preserved	Overall
Error (%)	IP	10.0 (2.3)	10.6 (2.5)	20.3 (3.3)	15.4 (3.3)	15.1 (2.1)	13.1 (2.0)	14.1 (1.5)
	Wd	13.0 (2.3)	9.9 (1.9)	14.6 (3.2)	10.4 (2.8)	13.8 (2.1)	10.2 (2.0)	12.0 (1.5)
	Overall (IP & Wd combined)	11.5 (1.6) **10.9** (1.2) (Del. & Pres. combined)	10.3 (1.6)	17.4 (2.3) **15.2** (1.7) (Del. & Pres. combined)	12.9 (2.2)	14.5 (1.5)	11.6 (1.4)	13.0 (1.1)
RT (ms)	IP	563 (37.0)	611 (36.8)	646 (29.4)	706 (40.0)	604 (32.1)	632 (34.9)	**618** (32.1)
	Wd	613 (29.4)	643 (32.1)	699 (36.4)	706 (40.0)	657 (31.7)	674 (34.5)	**665** (31.7)
	Overall (IP & Wd combined)	588 (23.6) **607** (22.9) (Del. & Pres. combined)	627 (24.4)	673 (25.9) **676** (25.4) (Del. & Pres. combined)	679 (28.9)	630 (22.5)	653 (24.5)	642 (22.5)

Comparisons which are statistically noteworthy as suggested by ANOVAs (C2 Identity × C2 Deletion × Prosodic Boundary) are highlighted in bold. (Numbers in parentheses are standard errors)

In ANOVAs on the errors, with the factors C2 Identity (/k/ vs. /p/), C2 Deletion (C2 deleted or preserved), and Prosodic Boundary (Word or IP boundary), there was a main effect only of C2 Identity which was significant by participants, but marginal by items: /k/ was detected, on average, 5% more accurately than /p/ ($F1(1, 91) = 4.262$, $p < 0.042$; $F2(1, 12) = 3.83$, $p < 0.08$). No other effects or interactions were significant. Critically, there was no C2 Deletion effect ($F1(1, 91) = 2.39$, $p > 0.1$; $F2(1, 12) = 1.16$, $p > 0.1$): The error rate in the C2-deleted condition was not significantly higher than that in the C2-preserved condition. Planned comparisons showed no significant differences between C2-deleted and C2-preserved conditions for each prosodic boundary and consonant type.

In ANOVAs on latencies, in line with error analyses, there was no main effect of C2 deletion ($F1(1, 91) = 2.67$, $p > 0.1$; $F2(1, 12) = 1.95$, $p > 0.1$) and an effect of C2 identity, with /k/s recognized on average 69 ms more quickly than /p/s ($F1(1, 91) = 14.98$, $p < 0.001$; $F2(1, 12) = 16.41$, $p < 0.005$). The C2 Identity factor did not interact with any other factors. Responses to /k/ were therefore both faster and more accurate than responses to /p/ across Prosodic Boundary and C2 Deletion conditions, as can be seen in Table 1. There was also a main effect of Prosodic Boundary significant by items and marginal by participants ($F1(1, 91) = 3.89$, $p < 0.08$; $F2(1, 12) = 4.59$, $p < 0.05$), which did not interact with any other factors: Targets were detected faster after IP than after Word boundaries.

In sum, listeners detected physically-absent targets (in the C2-deleted condition) as accurately and as fast as they detected physically-present targets (in the C2-preserved condition). They also generally performed better with /k/ than with /p/ (as reflected in both errors and latencies) and after IP versus Word boundaries (as reflected in latencies) regardless of whether C2 was deleted or preserved.

Summary and Discussion

The results of Experiment 1 showed that native speakers of Korean were as accurate and fast detecting physically-absent phonemes as detecting physically-present phonemes in consonant-cluster simplification environments. Because the physically-present target contains additional acoustic-phonetic cues, listeners were, all else being equal, expected to perform better in detecting preserved than deleted targets, but this was not the case. This suggests that something more than physical acoustic-phonetic cues was used by listeners in detecting the target. The results also showed a trend towards an effect of prosodic boundary: Listeners tended to be faster in detecting C2 targets when the target-bearing word was preceded by an IP boundary than when it was preceded by a Word boundary. This effect did not interact with C2 Deletion and C2 Identity factors, showing that the presence of an IP boundary facilitated the target detection of both deleted and preserved /k/ and /p/. Finally, there was also an effect of C2 Identity but crucially without any interaction with C2 Deletion: Listeners were generally more accurate and faster in detecting /k/ than /p/, regardless of whether the target was physically present or absent. The /k/-/p/ asymmetry therefore cannot be interpretable as coming from more robust acoustic-phonetic cues associated with physically-present /k/ versus /p/.

The most important of these findings is the fact that listeners were able to detect physically-absent (deleted) targets as efficiently as physically-present targets. This implies that the way in which listeners recover from continuous-speech production processes cannot be due entirely to language-universal perceptual processes which use low-level acoustic phonetic information (e.g., as in the feature-parsing account of assimilation proposed by Gow (2001, 2002), Gow and Im (2004)). Given that no acoustic information about the targets was available when they were deleted, language-specific knowledge must have been employed to recover their underlying representations. The most relevant language-specific

knowledge is phonological knowledge—i.e., the phonological phenomenon that in a tri-consonantal sequence C1C2C3 either C1 (/l/) or C2 (/k/ or /p/) is deleted. This knowledge may help listeners to restore the deleted phoneme C2 (e.g., via a process of phonological inference; Gaskell and Marslen-Wilson 1996, 1998).

As noted earlier, however, listeners can use contextual lexical-semantic information in lexical processing (e.g., Dahan and Tanenhaus 2004; Mattys et al. 2005; Tyler and Wessels 1983). In the stimuli of Experiment 1, the preceding contextual word was semantically related to the target-bearing word (e.g., $not^h\ddot{\imath}$ 'notebook' and *ilk-ta,* 'to read'). As suggested by Mattys et al. (2005), higher-order information (in this case, the contextual lexical-semantic information) may be weighted more than lower-order information (in this case, phonological information). It is therefore possible that the contextual support in Experiment 1 contributed to the recognition of the target-bearing word, and that this in turn reduced processing differences between C2-deleted and C2-preserved variants. If this were the case, the recovery of deleted phonemes could be due to listener's use of contextual lexical-semantic knowledge, rather than language-specific phonological knowledge. In Experiment 2, we further explore this question by examining how listeners' performance changes when the contextual support is removed from the input.

Experiment 2

In this experiment, we examine how listeners detect target phonemes (/p, k/ in /lpt/ or /lkt/) when the information in the preceding context was removed from the input by low-pass filtering. Of course, given the lack of contextual lexical-semantic support, listeners' performance, whether the target is physically present or absent, is expected to be poorer than their performance in the presence of such contextual support (in Experiment 1). Crucially, if the Experiment 1 listeners' ability to restore physically-absent phonemes stemmed primarily from the use of lexical knowledge associated with contextual information, listeners in Experiment 2, with the input devoid of contextual lexical-semantic support, should find it much harder to restore deleted phonemes than to detect physically-present phonemes. Alternatively, however, if language-specific phonological knowledge is still effectively used in restoring deleted phonemes, the Experiment 2 listeners should still be able to restore them efficiently, and perhaps even so efficiently that they perform as well on deleted targets as on physically-present targets.

Experiment 2 will also allow us to test whether the asymmetry between /k/ and /p/ found in Experiment 1 (overall better detection of /k/ than /p/) has to do with different degrees of contextual support. If preceding contexts happened to have stronger semantic association with /k/-bearing words than with /p/-bearing words, they could have helped listeners more with detection of /k/ than of /p/. If this were the case, the asymmetry is expected to disappear when the contextual semantic information is no longer available.

Method

Participants

An additional 96 students at Hanyang University, Seoul, were paid to participate. As in Experiment 1, they were all native Korean speakers, who had all been taught English as a second language at secondary school.

Materials and Procedure

The same materials and procedure as in Experiment 1 were employed, except that the contextual information preceding the target-bearing word was removed. The removal of the contextual information was carried out by applying a FFT (Hanning) low-pass filter to the preceding contextual word (with a cutoff frequency of 350 Hz and a 2048-point window size). The reduction in amplitude in the contextual word due to the filtering was compensated by increasing the filtered signals' amplitude up to the level of the original signals. The intelligibility of the filtered portion of each stimulus was initially checked by the first author, and confirmed by two more native speakers of Korean. They all reported that they could not identify the words in the filtered portion of each stimulus.

Results

As in Experiment 1, each subject was expected to provide at least two correct responses for either /p/ or /k/ in the C2-preserved condition. Two subjects failed this criterion. Table 2 shows mean error rates and latencies for the remaining 94 subjects.
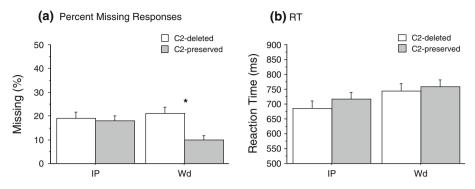
ANOVAs on the errors showed a main effect of C2 identity by participants, but not by items ($F1(1, 90) = 10.96$, $p < 0.001$; $F2(1, 12) = 1.299$, $p > 0.1$). This time, however, the trend was opposite to that in Experiment 1: /k/ was detected on average 6% *less* accurately than /p/. Again, contrary to the results of Experiment 1, there was a main effect of C2 deletion significant by participants and marginal by items ($F1(1, 90) = 9.42$, $p < 0.005$; $F2(1, 12) = 3.97$, $p < 0.08$): Responses in C2-preserved conditions were, on average, 6% more accurate than in C2-deleted conditions. Importantly, however, the C2 deletion effect interacted with Prosodic Boundary ($F1(1, 90) = 8.71$, $p < 0.01$; $F2(1, 12) = 4.63$, $p < 0.05$). Planned comparisons showed that error rates were significantly lower for the C2-preserved than for the C2-deleted condition in the Word boundary condition (by 11%, on average; $t1(91) = 3.823$, $p < 0.0001$; $t2(13) = 2.826$, $p < 0.05$), but not in IP boundary condition (a mean difference of only 1%; $t1(91) < 1$, $t2(13) < 1$). Thus, the physical presence of C2 indeed improved accuracy of listener's phoneme monitoring (i.e., it was more accurate relative to the C2-deleted condition) when the contextual information was removed, but only in the Word boundary condition. No other interactions (including C2 identity) were found, indicating that the effect was consistent for both /k/ and /p/. This effect is shown in Fig. 1a. (Further comparisons conducted separately for /k/ and /p/ confirmed the preserved-deleted effect in the Word boundary condition, significant by both participants and items (/k/, $t1(45) = 3.078$, $p < 0.01$; $t2(7) = 2.511$, $p < 0.05$; /p/, $t1(45) = 2.297$, $p < 0.05$; $t2(5) = 2.735$, $p < 0.05$)).

ANOVAs on latencies showed no main effect of C2 Identity ($F1(1, 90) < 1$, $F2(12) < 1$) and C2 Deletion ($F1(1, 90) = 2.304$, $p > 0.1$; $F2(1, 12) = 1.467$, $p > 0.1$). There was only a trend effect of Prosodic Boundary which was significant only by items ($F1(1, 90) = 1.637$, $p > 0.1$; $F2(12) = 7.987$, $p < 0.05$). The effect was not accompanied by an interaction with other factors. Responses to the target therefore tended to be faster in the IP than in the Word boundary condition (a mean difference of 50 ms) across C2 Identity and C2 Deletion conditions (see Fig. 1b) even when the contextual information was removed.

The only interaction found in the latency analyses was the one between C2 Identity and C2 deletion, but the effect was significant only by participants ($F1(1, 90) = 7.548$, $p < 0.01$; $F2(1, 12) = 2.170$, $p > 0.1$). This weak interaction stemmed from the tendency towards faster responses to the target in C2-deleted than in C2-preserved condition for /k/ (significant

Table 2 Mean proportion of errors (%) and mean Reaction Times (RT) in Experiment 2 (with low-pass filtered context), for phoneme monitoring responses to the targets /k/ and /p/ in underlyingly triconsonantal (C1C2C3) sequences

Measures	Prosodic boundary	C2 = /k/		C2 = /p/		Overall (/k/ & /p/ combined)		
		Deleted	Preserved	Deleted	Preserved	Deleted	Preserved	Overall
Error (%)	IP	20.8 (3.5)	23.4 (2.9)	17.4 (3.5)	12.5 (2.3)	19.1 (2.5)	18.0 (1.8)	18.5 (1.6)
	Wd	25.5 (2.8)	12.5 (2.7)	16.7 (3.5)	7.2 (2.3)	**21.1 (2.6)**	**9.9 (1.8)**	15.5 (1.7)
	Overall (IP & Wd combined)	23.2 (2.2)	18.0 (2.0)	17.0 (2.5)	9.9 (1.6)	20.1 (1.8)	13.9 (1.3)	17.0 (1.2)
		20.6 (1.2) (Del. & Pres. combined)		**13.4 (1.8)** (Del. & Pres. combined)				
RT (ms)	IP	680 (32.4)	734 (29.1)	693 (34.9)	698 (38.1)	686 (29.7)	716 (28.6)	**701 (27.2)**
	Wd	693 (28.8)	768 (31.3)	793 (40.8)	750 (30.7)	743 (30.4)	759 (29.3)	**751 (27.8)**
	Overall (IP & Wd combined)	**687 (21.7)**	**751 (21.3)**	**743 (26.8)**	**724 (24.6)**	715 (21.2)	738 (20.5)	726 (19.4)
		715 (21.2) (Del. & Pres. combined)		738 (20.5) (Del. & Pres. combined)				

Comparisons which are statistically noteworthy as suggested by ANOVAs (C2 Identity × C2 Deletion × Prosodic Boundary) are highlighted in bold. (Numbers in parentheses are standard errors)

Fig. 1 Experiment 2: percent missing responses (%) and reaction time (RT, in ms) by Prosodic Boundary (IP, Wd) and C2 Deletion (C2-deleted, C2-preserved). *Error bars* indicate standard errors and the *asterisk* refers to statistical significance both by participants and items in planned pairwise comparisons

only by participants: t1(93) = −2.783, p < 0.01; t2(15) = −1.470, p > 0.1), but not for /p/ (t1(93) < 1, t2(11) < 1; see means in Table 2). No other interactions were found.

Given that detection accuracy was lower in the C2-deleted condition (at least in the Word boundary condition), it was further tested whether this was due to a speed-accuracy trade-off. Pairwise comparisons separated by Prosodic Boundary and C2 Identity showed that faster responses to the deleted target were found only for /k/ in the Word boundary condition. This effect was again significant only by participants (for /k/, t1(45) = −2.366, p < 0.05, t2(7) = −1.046, p > 0.1). This suggests that the tendency towards less accurate responses to deleted /k/s in the Word boundary condition may be at least partially attributable to a speed-accuracy tradeoff.

Combined Analyses

Combined analyses of Experiments 1 and 2 showed a main effect of Context in both error and latency analyses—i.e., listeners were both less accurate and slower without context (Experiment 2) than with context (in Experiment 1) (errors: means, 17.0% vs. 13.1%, F1(1, 181) = 5.97, p < 0.02, F2(1, 12) = 3.71, p < 0.08; latencies: means, 726.1 ms vs. 641.6 ms, F1(1, 181) = 8.37, p < 0.005, F2(1, 12) = 48.27, p < 0.001). There was also a main effect of Prosodic Boundary in latencies, this time significant both by participants and by items (F1(1, 181) = 4.81, p < 0.05, F2(1, 12) = 17.8, p < 0.001) without interaction with Context, showing that listeners were generally faster in detecting targets after IP than after Word boundaries in both Experiments 1 and 2.

The context effect did not interact with C2 Deletion in either errors or latencies. There was however a trend effect of a three-way interaction between Context, C2 Deletion and Prosodic Boundary in error analyses significant only by participants (F1(1, 181) = 4.32, p < 0.05; F2(1, 12) = 1.29, p > 0.1). This three-way interaction was in line with the results of Experiment 2 in that error rates were lower for the C2-preserved than for the C2-deleted condition only in the Word boundary condition, which was significant both by participants and items in Experiment 2.

Another two-way interaction was observed between C2 Identity and Context in the latency analyses, which was significant by participants and marginal by items (F1(1, 181) = 4.88, p < 0.05, F(1, 12) = 3.96, p < 0.07). This again was in line with the results of Experiments 1 and 2 in that /k/ tended to be detected faster than /p/ in Experiment 1 but the opposite was true in Experiment 2. No other effects were found.

Summary and Discussion

The listener's performance with phoneme detection was generally poorer, as expected, when the contextual information was removed, showing that lack of contextual information was disadvantageous regardless of whether C2 was physically present or absent. From this we can infer that contextual information in Experiment 1 indeed helped listeners with target detection, but its contribution to detecting physically-absent phonemes appears to be just as good as its contribution to detecting physically-present phonemes. Crucially, listeners were able to restore deleted phonemes with as low as 20% missing responses to the targets even when they had no contextual lexical-semantic information. This confirms that listeners used language-specific knowledge other than contextually-driven lexical knowledge in restoring physically-absent targets.

The results of Experiment 2 and the combined analyses of Experiments 1 and 2 also showed some patterns that were different from the results of Experiment 1. First, the physical presence of C2 induced better accuracy in the listener's phoneme monitoring than in the C2-deleted condition. But this effect was limited to the Word boundary condition. When the contextual information was removed, listeners appear to have paid more attention to the physical acoustic-phonetic information at least in this condition. Second, /k/ was no longer detected better than /p/. Instead, if anything, the opposite was true: /p/ tended to be detected more accurately than /k/. The /k/ advantage observed in Experiment 1 can therefore be interpreted as having more to do with the preceding context than with C2 Identity. Third, there was a trend (significant only by participants) towards an interaction effect between C2 Identity and C2 Deletion: Listeners tended to detect /k/ faster in the absence of physical target (C2-deleted condition), but no such effect was observed with /p/. Posthoc analyses suggested that the faster responses to the deleted /k/ could be due to a speed-accuracy tradeoff. Finally, in line with Experiment 1, there was a tendency towards faster responses to the target in the IP condition than in the Word boundary condition. This suggests that even when the contextual information was removed, the presence of the larger prosodic (IP) boundary before the target-bearing word helped listeners to detect the target regardless of whether the target was physically present or absent.

General Discussion

In Experiment 1 (with contextual lexical-semantic information), we found that Korean listeners were able to detect physically-absent phonemes as accurately as physically-present phonemes. Listeners restored deleted phonemes efficiently enough to offset the disadvantage from the complete lack of acoustic-phonetic information about those phonemes (as confirmed in the Pretest). It was proposed that the listeners' ability to restore deleted phonemes may be attributable to their use of either contextual lexical-semantic knowledge (the preceding context word was semantically related to the target-bearing word (e.g., $not^h i$, 'notebook' for the target-bearing word *ilk-ta,* 'to read') or language-specific phonological knowledge (i.e., that C2 may be deleted in the tri-consonantal cluster simplification process).

In Experiment 2, we tested whether listeners would be able to restore deleted phonemes even when the contextual lexical semantic information was eliminated from the input. The results showed that even without contextual support, listeners restored deleted phonemes efficiently, that is, with only 20% of errors. From the fact that no acoustic-phonetic information about the targets was available when they were deleted, we can conclude that listeners must have used language-specific knowledge in recovering underlying representations of the

phonologically deleted phonemes. These results thus imply that, while fine-grained phonetic details may serve as language-universal cues to underlying representations of phonologically-altered sounds in some cases (e.g., assimilated sounds, Gow 2001, 2002; Gow and Im 2004), language-specific phonological knowledge (e.g., via a process of phonological inference; Gaskell and Marslen-Wilson 1996, 1998) does play an important role in recovering phono-logically-altered sounds in other cases. As Mitterer and McQueen (2009) argue, it makes good sense that different mechanisms are required for different kinds of continuous-speech phenomena because each phenomenon makes different demands on the listener.

How then is language-specific phonological knowledge employed in online continuous speech processing? Specifying the exact mechanism underlying this ability is beyond the scope of the present study, but at least two possibilities can be considered. One is that the knowledge is stored lexically. On this view, the use of phonological knowledge could further be modulated by the frequency of occurrence of different phonological variants (Connine 2004; Ranbom and Connine 2007). Recall that Cho and Kim (2009) reported that although both C1-deleted and C2-deleted variants are phonologically possible, C2-deleted variants occur more frequently (about 50.8%) than C1-deleted (C2-preserved) variants (17.8%). It is thus possible that different phonological variants of the target-bearing words (i.e., C2-deleted, C1-deleted, and no deletion variants) are stored in the mental lexicon, and that those represen-tations are frequency-coded. On this view, language-specific phonological knowledge would be brought to bear in recovery from consonant-cluster simplification through retrieval, in a frequency-sensitive way, of the underlying tri-consonantal cluster from the lexicon. Because C2-deleted variants occur frequently, retrieval of the underlying form when such a variant is heard can be rapid enough to match performance in retrieving the underlying form of C2-preserved variants, which are less frequent but which contain acoustic information about C2.

An alternative mechanism is prelexical inference (Gaskell and Marslen-Wilson 1996, 1998). On this view recovery is not based on word-specific knowledge, but rather on knowl-edge abstracted over words. If the prelexical processor knows that C2 in a Korean tri-con-sonantal cluster can be deleted, then it can fill in the missing sound. This inferential process could again be frequency sensitive (as suggested by Gaskell 2003 for inferences about place assimilation and by Mitterer and Ernestus (2006) and Mitterer and McQueen (2009) for infer-ences about [t] lenition): When a phonological variant occurs more frequently than another variant with the same underlying representation, the phonological inference made from the former could be stronger. But note that the inference in the Korean simplification case cannot be complete. In a /l/-C2-/t/ sequence, C2 could be either /k/ or /p/. Hence, if the basic recovery mechanism is prelexical, with restoration of two different stops, then final recognition of the deleted sound would still have to be based on recovery of lexical knowledge. This makes the account based on a lexical mechanism more parsimonious.

The frequency-sensitive use of phonological knowledge in spoken-word recognition has implications for language acquisition, especially in terms of how the frequency-coded phonological variants (i.e., more frequent C2-deleted variants versus less frequent C1-deleted variants) are acquired by children. It has been well documented in the literature on phonological acquisition that liquids (e.g., /l/) are acquired later than stops (e.g., Stoel-Gammon 1985; Vihman 1996). This can be accounted for by differential markedness—that is, liquids are more marked than other consonants (Jakobson 1968; Chomsky and Halle 1968). It then follows that at a developmental stage where a Korean child has learned the simplification process that either C1 or C2 can be deleted, but has not acquired production of the liquid /l/ yet, the child is mostly likely to simplify the tri-consonantal /l/-C2-/t/ sequence by deleting C1 (/l/), which deviates from the adult simplification pattern in which /l/ is deleted less frequently. If this were the case, though it is not attested, the earlier acquisition

of /l/-deleted variants (before the acquisition of C2-deleted variants) could be said to be driven by universally applicable markedness constraints (see Macken 1995; Fikkert 2007 for some discussions of the roles of markedness relations in phonological acquisition).

How then could children become adult-like in simplifying consonant clusters once /l/ is acquired, so that they no longer delete /l/ (C1) across the board, but delete C2 (/p/ or /k/) more frequently? One possibility is through statistical learning. Statistical learning has been proposed as one of the basic mechanisms in language acquisition (e.g., Gómez and Gerken 2000; Saffran 2001, 2003; Saffran and Wilson 2003). Although the role of statistical learning in language acquisition has often been examined in terms of sequence learning (e.g., about transitional probability of phonemic sequences within and across syllables and words), it has been extended to statistical learning of higher-level linguistic structures such as words and grammatical features (e.g., Saffran and Wilson 2003). It is then possible that the frequency-coded lexical properties of consonant cluster simplification patterns are learned through statistical learning mechanisms as the lexicon is expanded in the course of language acquisition.

Other important findings of the present study are the effects of Prosodic Boundary (in Experiments 1 and 2) and its interaction with C2 Deletion in Experiment 2. Both Experiments 1 and 2 showed a tendency that listeners processed simplified forms (both C2-preserved and C2-deleted) better when the target-bearing words were preceded by a large prosodic (IP) boundary. Better target detection after IP boundaries could be attributable to facilitated lexical segmentation at the IP boundary, as suggested by previous studies (Christophe et al. 2004; Cho et al. 2007; Kim and Cho 2009)—that is, the presence of a phrasal prosodic boundary (e.g., a Phonological or IP boundary) facilitates lexical segmentation across that boundary.

Crucially, in Experiment 2, there was an interaction between Prosodic Boundary and C2 Deletion, showing that listeners detected deleted targets as accurately as preserved targets after IP boundaries. That is, in the context of an IP boundary there was no benefit from the presence of acoustic-phonetic information in the C2-preserved condition. However, the benefit of the acoustic-phonetic information in the physically-present target was present in the Word boundary condition—that is, physically-present targets were detected more accurately than physically-absent targets after Word boundaries. But it should be noted that even if the detection accuracy for deleted phonemes were significantly lower than for preserved phonemes, the missing responses to deleted phonemes in the Word condition were still as low as 21.1%, suggesting that listeners were able to recover the deleted phonemes quite efficiently. The question thus arises as to why lack of available acoustic-phonetic information with deleted C2 gave rise to poorer performance only after Word boundaries, and only with no contextual lexical-semantic support (i.e., only in Experiment 2). One possible explanation has to do with how much information is available in the speech input. In Experiment 1, it is possible that phonological knowledge in combination of contextual support could be sufficient in helping listeners restore deleted phonemes, and this appears to be enough to match the benefit of segmental acoustic-phonetic information for the preserved C2. In Experiment 2, with recourse to phonological knowledge alone, listeners still recovered the deleted targets efficiently. In addition, given that detection accuracy was no different between the C2-deleted and C2-preserved conditions with IP boundaries, it is reasonable to assume that IP boundary information, along with use of phonological knowledge, helped listeners with restoration of deleted phonemes. Exactly how IP boundary information helped with recovery of deleted targets is not clear, but facilitated lexical segmentation and lexical access at a phrase boundary (Christophe et al. 2004; Cho et al. 2007; Kim and Cho 2009) could help recover underlying representations efficiently. In contrast, the Word boundary condition in Experiment 2 contains neither contextual nor IP-boundary information. In this case, listeners have only

language-specific phonological knowledge to rely on. This may have led to relatively poorer performance detecting deleted targets in this condition.

The present study also has implications for the listener-oriented production hypothesis (Jun 2004). According to this hypothesis, speakers help listeners by exerting greater effort to preserve segments which have intrinsically more powerful acoustic cues. Although this hypothesis might explain speech production phenomena such as the strong cross-linguistic tendency for velars to be less subject to phonological alterations than labials or alveolars (Cho 1999; Jun 2004), and how listeners process unassimilated versus illegally assimilated /k/ versus /p/ (Cho and McQueen 2008), the present study does not provide further support for the hypothesis. In both Experiments 1 and 2, listeners showed no performance differences between /k/ and /p/ when they were physically present, suggesting that listeners do not always benefit more from the presence of /k/ versus /p/.

The discrepancy between the results of Cho and McQueen (2008) and the present study may have stemmed from the phonological viability of the materials tested in the two studies. Cho and McQueen (2008) created an illegal assimilation case in which /k/ and /p/ were illegally assimilated to the following /t/, and compared listener's detection of unaltered (unassimilated) /k/ and /p/ with the illegally assimilated ones (both realized as [t]). So, in their study, recovery of the underlying representations of illegally assimilated sounds did not involve use of phonological knowledge. In the present study, however, the present and absent targets were both phonologically viable, and therefore phonological knowledge was most likely referred to in recovery of deleted phonemes. It is therefore hard to determine how much the use of phonological knowledge may have masked the effects of acoustic-phonetic differences between /k/ versus /p/. Nevertheless, given that there were no other differences between /k/ and /p/, and that contextual information was removed from the stimuli in Experiment 2, the present null effect does suggest that differences in perceptual robustness of different sounds, if they exist, are not always used in speech processing.

Finally, two other issues should be discussed: the possible role of orthographic knowledge in consonant-cluster simplification, and whether there are possible effects of second-language experience on how Korean listeners recover deleted consonants. Regarding possible orthographic effects, it should first be noted that all our participants were fully literate university students, and that literacy was not a factor we investigated. Strong conclusions about effects of literacy are thus not warranted. We would nevertheless like to suggest that learning to read is unlikely to have fundamentally changed how Korean listeners recover from consonant-cluster simplification. First, Korean children have to be able to recognise spoken Korean words with deleted consonants before they learn to read. Basic mechanisms (such as frequency-sensitive lexical storage of pronunciation variants, see above) should thus have been acquired prior to acquisition of literacy. Second, the Korean writing system (Hangul) is a fully transparent alphabetic script. Although learning such a script is likely to strengthen phonological representations, it is unlikely to change those representations qualitatively or thus alter the nature of phonological processing (unlike, e.g. in English, where learning the irregularities of sound-print correspondences could result in changes in how spoken words are recognized, see, e.g., Taft et al. 2008). We suggest, however, that learning to read may well have made it easier for our participants to perform the phoneme monitoring task that we used. The acquisition of orthographic knowledge, especially when it is alphabetic as in English or in Korean, has a significant impact on phonological awareness, and hence on participants' ability to perform metaphonological tasks (e.g., Mann 1986; Morais 1985; Morais et al. 1986; Read et al. 1986; Taft 1991, 2006; Ziegler and Ferrand 1998; Ziegler and Muneaux 2007). Learning to read thus probably made it easier for the participants in our study to reflect on the phonological structure and content of the consonant clusters they heard, and hence to do

the phoneme monitoring task, but probably did not change the representations of words with tri-consonantal sequences.

Let us now consider possible effects of second-language experience. It has been well documented in the literature that speech perception and production are not the same in monolinguals and bilinguals (e.g., Tees and Werker 1984; Gottfried 1984; Flege and Hillenbrand 1986; Mack 1989; Best and Strange 1992; Flege 1992; Munro 1993; Sancier and Fowler 1997; Bosch et al. 2000; Fowler et al. 2008). The general consensus is that highly proficient second-language learners or even simultaneous bilinguals do not attain the same level of competence in their second language (or either one of the two languages for simultaneous bilinguals) as monolingual speakers of that language. But because the current experiment was performed in the listeners' first language, one might expect no effects of second-language experience. Such effects are nonetheless possible, given that bilinguals may not use the same cues as pure monolinguals in speech processing (e.g., Mack 1989; Bosch et al. 2000). As we noted earlier, our subjects were all bilingual in the sense that they all started to learn English in secondary school (approximately when they were 12–13 years old). Although they were all college students at the time of their participation, their English proficiencies can be expected to vary. One could therefore predict that the recovery of deleted phonemes may also differ depending on English proficiency. Given that English allows tri-consonantal sequence as in 'pulped' and 'bulked,' more proficient Korean learners of English may be less sensitive to the phonotactic constraint in Korean which disfavors tri-consonantal sequences. One could then hypothesize that listeners' performance in processing simplified variants of tri-consonant sequences in Korean becomes poorer as their English proficiency increases. While the present study was not designed to test this hypothesis, it warrants further work. It would be interesting to know whether recovery of deleted phonemes in Korean is modulated by listeners' proficiency in English.

Conclusion

It is striking that Koreans can perceptually restore missing phonemes so efficiently. Kemps et al. (2004; see also Kemps 2004) showed that Dutch listeners, in phoneme monitoring, could restore segments that were missing from suffixes in reduced word-forms, but that their performance was poorer than when the segments were physically absent. We have shown here, however, that if the simplification process is frequent enough, restoration of a missing segment is as easy as detection of a physically-present segment, even when contextual support is not available. The recovery process is often efficient enough that monitoring performance on physically-absent targets can match performance on physically-present targets. The benefit of acoustic-phonetic information was found only when no other higher-order information (contextually-driven lexical support and phrase boundary information) was available. This suggests that listeners make differential use of different levels of information, depending on how much information is available in the speech input.

Most importantly, the efficient recovery of deleted phonemes without residual acoustic phonetic cues to deleted phonemes suggests that language-specific phonological knowledge can be used without recourse to language-universal perceptual processes which rely on low-level fine-grained phonetic detail (Gow and Im 2004; Mitterer et al. 2006). Further research is required to specify the mechanism underlying this ability. Our results nonetheless impose strong constraints on possible accounts: Listeners use language-specific phonological

knowledge when perceiving the results of continuous-speech production processes which leave no acoustic trace of underlying sounds.

Appendix

See Table 3.

Table 3 Experimental items

/k/-bearing two-word phrases	/p/-bearing two-word phrases
1. koŋki-ka malk-təlato	*1. kutu-lo palp-təlato*
'although the air is clear'	'although stepping (on it) with the shoes'
2. notʰi-lɨl ilk-təlato	*2. tʃatu-ka t*əlp-təlato*
'although reading the notebook'	'although the plum is astringent (unripe)'
3. kituŋ-i kulk-təlato	*3. sija-ka nəlp-təlato*
'although the pillar is thick'	'although the view is wide'
4. tʃomjəŋ-i palk-təlato	*4. tuk*e-ka jalp-təlato*
'although the light is bright'	'although the thickness is thin'
5. kaljəwə kɨlk-təlato	*5. tʃʰima-ka tʃ*alp-təlato*
'although scratching, because it's itching'	'although the skirt is short'
6. əntʃenka nɨlk-təlato	*6. pitʃʰk*al-i jəlp-təlato*
'although (we) get old one day'	'although the color is light (pale)'
7. noŋto-ka mulk-təlato	
'although the density is diluted (watery)'	
8. noɨl-i pulk-təlato	
'although the sky (at the sunset) is red'	

The Korean phrases are given as phonemic IPA transcriptions. '-' refers to a word-internal bound morpheme boundary. The diacritic * refers to the tense (fortis) stop in Korean (see Cho et al. 2002, for phonetic properties of obstruent consonants in Korean)

References

Best, C. T., & Strange, W. (1992). Effects of language-specific phonological and phonetic factors on cross-language perception of approximants. *Journal of Phonetics, 20*, 305–330.

Bosch, L., Costa, A., & Sebastián-Gallés, N. (2000). First and second language vowel perception in early bilinguals. *European Journal of Cognitive Psychology, 12*, 189–221.

Cho, T. (1999). Intra-dialectal variation in Korean consonant cluster simplification: A stochastic approach. *Chicago Linguistics Society, 35*(CLS 35), 43–57. (Also appeared in *UCLA working papers in linguistics, vol. 2, papers in phonology 3*, 207–226, edited by M. Gordon).

Cho, T., Jun, S.-A., & Ladefoged, P. (2002). Acoustic and aerodynamic correlates of Korean stops and fricatives. *Journal of Phonetics, 30*, 193–228.

Cho, T., & Keating, P. (2001). Articulatory and acoustic studies of domain-initial strengthening in Korean. *Journal of Phonetics, 29*, 155–190.

Cho, T., & Kim, S. (2009). Statistical patterns in consonant cluster simplificaiotn in Seoul Korean: Within-dialect interspeaker and intraspeaker variation. *Phonetics and Speech Sciences (Journal of the Korean Society of Speech Sciences), 1*, 33–40.

Cho, T., & McQueen, J. M. (2006). Phonological versus phonetic cues in native and nonnative listening: Korean and Dutch listeners' perception of Dutch and English consonants. *Journal of the Acoustical Society of America, 119*, 3085–3096.

Cho, T., & McQueen, J. M. (2008). Not all assimilated sounds are perceived equally evidence from Korean. *Journal of Phonetics, 36*, 239–249.

Cho, T., McQueen, J. M., & Cox, E. (2007). Prosodically driven phonetic detail in speech processing: The case of domain-initial strengthening in English. *Journal of Phonetics, 35*, 210–243.

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

Christophe, A., Peperkamp, S., Pallier, C., Block, E., & Mehler, J. (2004). Phonological phrase boundaries constrain lexical access I. Adult data. *Journal of Memory and Language, 51*, 523–547.

Coenen, E., Zwitserlood, P., & Boelte, J. (2001). Variation and assimilation in German: Consequences for lexical access and representation. *Language and Cognitive Processes, 16*, 535–564.

Connine, C. M. (2004). It's not what you hear but how often you hear it: On the neglected role of phonological variant frequency in auditory word recognition. *Psychonomic Bulletin & Review, 11*, 1084–1089.

Dahan, D., & Tanenhaus, M. K. (2004). Continuous mapping from sound to meaning in spoken-language comprehension: Immediate effects of verb-based thematic constraints. *Journal of Experimental Psychology: Learning, Memory and Cognition, 30*, 489–513.

Edwards, J., Beckman, M. E., & Fletcher, J. (1991). Articulatory kinematics of final lengthening. *Journal of the Acoustical Society of America, 89*, 369–382.

Fikkert, P. (2007). Acquiring phonology. In P. de Lacy, *Handbook of phonological theory* (pp. 537–554). Cambridge, MA: Cambridge University Press.

Flege, J. E. (1992). Speech learning in a second language. In C. A. Ferguson, L. Menn, & C. Stoel-Gammon (Eds.), *Phonological development: Models, research and implications* (pp. 565–604). Timonium, MD: York Press.

Flege, J. E., & Hillenbrand, J. (1986). Differential use of temporal cues to the /s/-/z/ contrast by native and non-native speakers of English. *Journal of the Acoustical Society of America, 79*, 508–517.

Flemming, E. (1995). *Auditory features in phonology*. Ph.D. dissertation, UCLA, Los Angeles, CA.

Fowler, C. A., Sramko, V., Ostry, D. J., Rowland, S. A., & Halle, P. (2008). Cross language phonetic influences on the speech of French–English bilinguals. *Journal of Phonetics, 36*, 649–663.

Gaskell, M. G. (2003). Modelling regressive and progressive effects of assimilation in speech perception. *Journal of Phonetics, 31*, 447–463.

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.

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.

Gómez, R. L., & Gerken, L. (2000). Infant artificial language learning and language acquisition. *Trends in Cognitive Sciences, 4*, 178–186.

Gottfried, T. L. (1984). Effects of consonant context on the perception of French vowels. *Journal of Phonetics, 12*, 91–114.

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

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

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

Hayes, B., Kirchner, R., & Steriade, D. (Eds.). (2004). *Phonetically based phonology*. Cambridge: Cambridge University Press.

Hume, E. & Johnson, K. (Eds.). (2001). *The role of speech perception in phonology*. San Diego, CA: Academic Press.

Jakobson, R. (1968). *Child language, aphasia, and phonological universals*. The Hague: Mouton.

Jun, J. (1995). *Perceptual and articulatory factors in place assimilation: An optimality theoretic approach*. Ph.D. dissertation, UCLA.

Jun, J. (2004). Place assimilation. In B. Hayes, R. Kirchner, & D. Steriade (Eds.), *Phonetically based phonology* (pp. 58–86). Cambridge: Cambridge University Press.

Jun, S.-A. (1993). *The phonetics and phonology of Korean prosody*. Ph.D. dissertation, The Ohio State University.

Jun, S.-A. (2000). K-ToBI (Korean ToBI) labelling conventions version 3. *Speech Sciences (Journal of Korean Association of Speech Sciences), 7*, 143–169. (Also published in *UCLA Working Papers in Phonetics, 99*, 149–173).

Kemps, R. (2004). *Morphology in auditory lexical processing: Sensitivity to fine phonetic detail and insensitivity to suffix reduction*. Ph.D. dissertation, Radboud University, Nijmegen (MPI Series in Psycholinguistics, 26).

Kemps, R., Ernestus, M., Schreuder, R., & Baayen, H. (2004). Processing reduced word forms: The suffix restoration effect. *Brain and Language, 90*, 117–127.

Kim-Renaud, Y.-K. (1976). *Korean consonantal phonology*. Ph.D. dissertation, University of Hawaii.

Kim, S. (2003). The role of post-lexical tonal contours in word segmentation. In M. J. Solé, D. Recasens, & J. Romero (Eds.), *Proceedings of the 15th international congress of phonetic sciences*. Adelaide Causal Productions.

Kim, S. (2004). *The role of prosodic phrasing in Korean word segmentation*. Ph.D. dissertation, UCLA.

Kim, S., & Cho, T. (2009). The use of phrase-level prosodic information in lexical segmentation: Evidence from word-spotting experiments in Korean. *Journal of the Acoustical Society of America, 125*, 3373–3386.

Kohler, K. J. (2000). Investigating unscripted speech implications for phonetics and phonology. *Phonetica, 57*, 85–94.

Mack, M. (1989). Consonant and vowel perception and production: Early English-French bilinguals and English monolinguals. *Perception & Psychophisics, 46*, 187–200.

Macken, M. (1995). Phonological acquisition. In J. A. Goldsmith (Ed.), *The handbook of phonological theory* (pp. 671–696). Cambridge, MA: Blackwell.

Mann, V. A. (1986). Phonological awareness: The role of reading experience. *Cognition, 24*, 65–92.

Mattys, S. L., White, L., & Melhorn, J. F. (2005). Integration of multiple speech segmentation cues: A hierarchical framework. *Journal of Experimental Psychology: General, 134*, 477–500.

McQueen, J. M. (2007). Eight questions about spoken-word recognition. In G. Gaskell (Ed.), *The Oxford handbook of psycholinguistics* (pp. 37–53). Oxford: Oxford University Press.

Mitterer, H., & Ernestus, M. (2006). Listeners recover /t/'s that speakers reduce evidence from /t/-lenition in Dutch. *Journal of Phonetics, 34*, 73–103.

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.

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 & Performance, 35*, 244–263.

Morais, J. (1985). Literacy and awareness of the units of speech: Implications for research on the units of perception. *Linguistics, 23*, 707–721.

Morais, J., Bartelson, P., Cary, L., & Alegria, J. (1986). Literacy training and speech segmentation. *Cognition, 24*, 45–64.

Munro, M. (1993). Non-native productions of English vowels: Acoustic measurements and accentedness ratings. *Applied Psycholinguistics, 36*, 39–66.

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.

Ranbom, J. L., & Connine, C. M. (2007). Lexical representation of phonological variation in spoken word recognition. *Journal of Memory and Language, 57*, 273–298.

Read, C. A., Zhang, Y., Nie, H., & Ding, B. (1986). The ability to manipulate speech sounds depends on knowing alphabetic reading. *Cognition, 24*, 31–44.

Saffran, J. R. (2001). Words in a sea of sounds: The output of infant statistical learning. *Cognition, 81*, 149–169.

Saffran, J. R. (2003). Statistical language learning: Mechanisms and constraints. *Current Directions in Psychological Science, 12*, 110–114.

Saffran, J. R., & Wilson, D. P. (2003). From syllables to syntax: Multilevel statistical learning by 12-month-old infants. *Infancy, 4*, 273–284.

Sancier, M., & Fowler, C. A. (1997). Gestural drift in a bilingual speaker of Brazilian Portuguese and English. *Journal of Phonetics, 25*, 421–436.

Spinelli, E., McQueen, J. M., & Cutler, A. (2003). Processing resyllabified words in French. *Journal of Memory and Language, 48*, 233–254.

Steriade, D. (1999). Phonetics in phonology the case of laryngeal neutralization. In M. Gordon (Ed.), *UCLA working papers in linguistics no. 2 papers in phonology 3* (pp. 25–146).

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

Stoel-Gammon, C. (1985). Phonetic inventories, 15–22 months: A longitudinal study. *Journal of Speech and Hearing Research, 28*, 505–512.

Taft, M. (1991). *Reading and the mental lexicon*. Hove, UK: Erlbaum.

Taft, M. (2006). Orthographically influenced abstract phonological representation: Evidence from non-rhotic speakers. *Journal of Psycholinguistic Research, 35*, 67–78.

Taft, M., Castles, A., Davis, C., Lazendic, G., & Nguyen-Hoan, M. (2008). Automatic activation of orthography in spoken word recognition: Pseudohomograph priming. *Journal of Memory and Language, 58*, 366–379.

Tees, R. C., & Werker, J. F. (1984). Perceptual flexibility: Maintenance or recovery of ability to discriminate non-native speech sounds. *Canadian Journal of Psychology, 38*, 579–590.

Tyler, L. K., & Wessels, J. (1983). Quantifying contextual contributions to word-recognition processes. *Perception & Psychophysics, 34*, 409–420.

van Donselaar, W., Kuijpers, C., & Cutler, A. (1999). Facilitatory effects of vowel epenthesis on word processing in Dutch. *Journal of Memory and Language, 41*, 59–77.

Vihman, M. M. (1996). *Phonological development: The origins of language in the child*. Oxford: Blackwell.

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

Wightman, C., Shattuck-Hufnagel, S., Ostendorf, M., & Price, P. (1992). Segmental durations in the vicinity of prosodic phrase boundaries. *Journal of the Acoustical Society of America, 91*, 1707–1717.

Ziegler, J. C., & Ferrand, L. (1998). Orthography shapes the perception of speech: The consistency effect in auditory word recognition. *Psychonomic Bulletin & Review, 5*, 683–689.

Ziegler, J. C., & Muneaux, M. (2007). Orthographic facilitation and phonological inhibition in spoken word recognition: A developmental study. *Psychonomic Bulletin & Review, 14*, 75–80.