Phonological versus phonetic cues in native and non-native listening: Korean and Dutch listeners' perception of Dutch and English consonants

Taehong Cho[a]
Hanyang University, Division of English Language and Literature, Seoul (133-791), Korea and Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands

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

We investigated how listeners of two unrelated languages, Korean and Dutch, process phonologically viable and nonviable consonants spoken in Dutch and American English. To Korean listeners, *released* final stops are nonviable because word-final stops in Korean are never released in words spoken in isolation, but to Dutch listeners, *unreleased* word-final stops are nonviable because word-final stops in Dutch are generally released in words spoken in isolation. Two phoneme monitoring experiments showed a phonological effect on both Dutch and English stimuli: Korean listeners detected the unreleased stops more rapidly whereas Dutch listeners detected the released stops more rapidly and/or more accurately. The Koreans, however, detected released stops more accurately than unreleased stops, but only in the non-native language they were familiar with (English). The results suggest that, in non-native speech perception, phonological legitimacy in the native language can be more important than the richness of phonetic information, though familiarity with phonetic detail in the non-native language can also improve listening performance. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2188917]

I. INTRODUCTION

In the first months of life, infants are typically able to discriminate acoustic-phonetic differences irrespective of whether the phonetic differences are phonologically contrastive in the ambient language or not (see, e.g., Jusczyk, 1997, for a review). By the end of their first year, however, their sensitivity to speech sound contrasts is generally refined according to the phonological properties of the ambient, native language (e.g., Werker *et al.*, 1981; Werker and Lalonde, 1988; see Cutler and Broersma, 2005, for a review). Although the timing of these refinements varies (Polka *et al.*, 2001), the general developmental pattern is that sensitivity to phonetic differences which are phonologically noncontrastive in the native language becomes attenuated relative to sensitivity to contrastive phonetic differences, and that the degree of such an attenuation is typically correlated with age (e.g., Flege, 1995). Eventually, perception of speech sounds by adults is highly biased or modulated by their experience with the phonological systems of their native language (there are effects, e.g., of phoneme repertoire, phonotactics, allophonic distribution, rhythmic structure and morphophonemic alternations; see e.g., Best, 1995; Best and Strange, 1992; Broersma, 2005; Costa *et al.*, 1998; Cutler *et al.*, 1986; Cutler and Otake, 1994; Flege, 1995; Flege and Hillenbrand, 1986; Hallé *et al.*, 1999; Kuhl and Iverson, 1995; Otake *et al.*, 1993; Weber, 2001; Weber and Cutler, 2006). Adult native Japanese listeners, for example, generally perceive both contrastive English sounds /r/ and /l/ as the flap /ɾ/, which is the sound in their native language that is phonetically closest to the two sounds (e.g., Best *et al.*, 1988; Best and Strange, 1992; see also Best, 1995). This is a case in which two contrastive categories in the non-native language are "assimilated" to the same phonological (or phonetic) category in the listener's native language.

Another example of native-language influences comes from differential use of phonetic cues by native and non-native listeners in the perception of contrastive sounds. Flege and Hillenbrand (1986), for instance, showed that in identifying fricatives as /s/ or /z/ (as in *peace* and *peas*) in American English, native English listeners used two well-known phonetic cues to the syllable-final voicing contrast, that is, fricative duration (which is longer for /s/ than for /z/) and preceding vowel duration (which is shorter before /s/ than before /z/; e.g., Denes, 1955, Raphael, 1972; see Watson, 1983 for a review). Non-native listeners of English, specifically, Swedish and Finnish listeners, who have no phonemically contrastive /s/-/z/ pair in their native language, but do have contrastive long and short vowels in their native language, used only vowel length differences to differentiate /s/ from /z/. According to Flege and Hillenbrand, listeners of Swedish and Finnish might have "reinterpreted" the role of phonologically contrastive vowel duration in their native language as a cue to the voicing contrast in non-native listening. Dutch listeners, however, appear not to use vowel duration in identification of voiced versus voiceless English syllable-final fricatives (Broersma, 2005). This may be because, in Dutch, /s/ and /z/ are phonemically contrastive, but not in

---

[a]Electronic mail: tcho@hanyang.ac.kr

syllable final position. Dutch listeners may therefore have learned to use cues other than vowel duration in perceiving the fricative voicing distinction.

What has emerged from these studies on native versus non-native listening is that speech perception is tuned according to language experience, resulting in a "phonological reorganization of speech sound percepts" (Cutler and Broersma, 2005), and that one of the driving forces behind this phonological reorganization is the various language-specific constraints that arise within the phonological system of the listener's native language. It is thus not surprising, as Cutler and Broersma (2005) noted, that improving performance of adult listeners in the perception of non-native sounds has been found to be very difficult although training can improve performance to some extent (e.g., Logan *et al.*, 1991; Lively *et al.*, 1993, 1994; Bradlow *et al.*, 1997).

In the present study, we further explore how listening is tuned by language experience. We investigate how listeners of two unrelated languages, Korean and Dutch, process sounds which are phonologically viable or nonviable in their native language. In Korean, stops in syllable- or word-initial position are lenis, fortis, or aspirated, and their phonetic cues are generally present in the release portion and at the beginning of the following vowel (Cho *et al.*, 2002; Kim *et al.*, 2002). In coda position, however, stops are unreleased, at least when the syllable or word is produced in isolation. Coda stops in Korean can be realized either with or without audible releases. When the coda stop is followed by a vowel-initial morpheme (due to morpheme concatenation) or by a vowel-initial word (due to prosodic grouping), it is resyllabified as the onset of the following syllable, and is released with an audible release noise. When followed by another stop, however, coda stops can be released with only an inaudible brief transient noise. Kim and Jongman (1996) reported that when underlying stem-final /t/ was followed by the velar stop /k/ in Korean, it was released with a brief transient 83% of the time. These events were difficult to perceive, but appeared in acoustic analyses. Henderson and Repp (1982) observed similar patterns for English stop sequences, and argued that what are traditionally known as "unreleased" stops are often produced with an inaudible articulatory release gesture. Korean coda stops followed by another stop thus tend, in this sense, to be "unreleased." Korean coda stops in isolated syllables (or utterance finally), however, are unreleased both auditorily and articulatorily. This means that the three-way manner contrast found in onset position in Korean is completely neutralized, without even subtle acoustic differences distinguishing among the three types of stop (Kim and Jongman, 1996). This behavior is often considered to be the outcome of an obligatory phonological rule which states that stops are unreleased when followed by a word boundary or by another obstruent within the word (e.g., Kim-Renaud, 1976; Martin, 1992).

In Dutch, however, voiceless coda stops in word-final position are generally released when words are produced in isolation (cf., Warner *et al.*, 2004; Ernestus and Baayen, in press). Like Korean, Dutch has released stops in syllable-initial (prevocalic) positions. The question we address here, therefore, is whether and how the position-specific phono-

logical difference between the two languages influences speech perception. Released coda stops in isolated syllables are phonologically nonviable in Korean; unreleased stops in this environment are phonologically nonviable in Dutch. It is important to emphasize that we use the term "phonological" to refer to language-specific sound patterns which are driven by obligatory phonological rules. The term "phonological knowledge" is thus used to refer to the native listener's knowledge not only about the native language's phoneme system but also about sound patterns determined by phonological rules. If the way native listeners perceive non-native speech sounds is guided by their native phonological knowledge, listeners should find it harder to process non-native sounds that deviate from native language phonological constraints (phonologically nonviable sounds) than to recognize those that are in conformity with those native constraints (phonologically viable sounds). This will be referred to as the *phonological-superiority hypothesis*.

There is a competing hypothesis, however. Released stops are produced with two distinct sets of acoustic-phonetic cues (those associated with the VC formant transition and those associated with the release burst) whereas unreleased stops are cued solely by the VC formant transition information. It has widely been agreed that the release burst is one of the most important cues to the identity of a stop, such that the presence of an audible release burst (in addition to formant transitions) increases the perceptual recoverability of the stop consonant (e.g., Stevens and Blumstein, 1978; Mattingly, 1981; Silverman, 1995; Smits *et al.*, 1996; Steriade, 1997; Kochetov, 2001; Zsiga, 2003; Wright, 2004; Flemming, 2005). That is, the information available to the listener is richer with than without the release burst.

In a perceptual study on Russian stops in syllable- or word-final coda positions, for example, Kochetov (2001) has shown that the identification of plain and palatalized coronal coda stops in Russian was significantly faster in the presence than in the absence of an audible release burst. With respect to stops in syllable-onset positions, the data reported by Smits *et al.* (1996) showed that in classifications of Dutch stops (/p/ vs /t/ vs /k/), Dutch listeners relied more on the release bursts than they did on the Consonant-Vowel (CV) formant transitions, when the stimuli contained a mismatch between the release burst and the CV formant transition. Similarly, Stevens and Blumstein (1978) showed that the presence of noise bursts at stop onset enhanced overall consonant classification in English. The perceptual importance of the release burst as a cue to stop identity has also been used as phonetic grounds for cross-linguistic phonological patterns of place and laryngeal contrasts for stop consonants (e.g., Steriade, 1997; Silverman, 1995; Wright, 2004; Flemming, 2005). Across languages, phonological place and laryngeal contrasts occur far more often in positions where the release bursts are available to cue the contrasts.

There is therefore considerable phonetic and phonological evidence for the importance of the release burst as a cue to stop identity. Thus, from the perspective of the relative richness of acoustic-phonetic cues, one might expect that released stops (with bursts) are processed more efficiently than

unreleased stops (with no bursts) irrespective of the phonological system of the listener's native language. This will be called the *phonetic-superiority hypothesis*.

In this study, we tested the *phonological-superiority* hypothesis against the *phonetic-superiority* hypothesis, by conducting two phoneme monitoring experiments (Connine and Titone, 1996; Otake *et al.*, 1996). Experiment 1 measured Korean listening performance; Experiment 2 examined Dutch listeners' performance on the same materials. Both groups of listeners heard four sets of speech materials: released and unreleased stops in Dutch, and released and unreleased stops in English. Both Dutch and English speech materials were included in order to examine whether the nature of the input language influences phoneme recognition. In Experiment 1, the language factor was whether the non-native speech materials were familiar or unfamiliar (to Korean listeners). In Experiment 2, it was whether the speech materials were native or non-native (to Dutch listeners).

II. EXPERIMENT 1

Korean listeners were presented with Dutch spoken stimuli in Experiment 1a and with English spoken stimuli in Experiment 1b. The listeners were asked to detect both nasal and oral consonants, in VC syllables. The oral consonants were either released or unreleased stops. We asked how speed and accuracy in phoneme monitoring varies as a function of the phonological viability of the stimuli in the listener's native language and as a function of the phonetic richness of those stimuli. On the one hand, the phonological-superiority hypothesis predicts that Korean listeners will detect unreleased stops more easily than released stops (because word-final stops in words spoken in isolation are always unreleased in Korean). On the other hand, the phonetic-superiority hypothesis predicts that Korean listeners will detect released final stops more easily than unreleased (because released stops are phonetically richer). The nasal targets, which were identical in the released and unreleased conditions, were included partly as distractors. Their inclusion also allowed us to examine whether phoneme monitoring performance on nasal targets varied depending on whether they co-occurred with released or unreleased stops. Finally, we tested whether phoneme monitoring performance is constrained by language experience (unfamiliar versus familiar speech stimuli, i.e., Dutch versus English).

A. Method

1. Participants

Fifty-two Korean student volunteers at Korea University in Seoul were paid to take part. They were divided into four groups of 13 according to the release condition (Released versus Unreleased) and the language condition (Dutch in Experiment 1a and English in Experiment 1b).

2. Materials

Two sets of materials were constructed in each language, one consisting of VC syllables with the release intact (the released condition), and one with the release spliced out (the unreleased condition). Within each set of VC's, there were 30 experimental items with 3 oral (/p, t, k/) and 3 nasal (/m, n, ŋ/) target consonants in their codas in 5 different vowel contexts (/a,e,i,o,u/; 6 target consonants × 5 vowel contexts =30 experimental items). Lists of four-syllable sequences were then constructed for each target-bearing experimental item with the target placed in either the last or the penultimate syllable, as in [V, V, V, VC] or [V, V, VC, V]. (Note that in the unreleased condition "VC" refers to a stimulus in which the released portion was spliced out, and therefore phonetic cues to "C" were present only in the preceding vowel.) Single V fillers (non-target-bearing syllables) were used instead of VC fillers in order to control for the frequency of occurrence of the target consonants. All V's in each list were different (e.g., [e, u, ak, i]). In addition to the 30 experimental lists, 15 nonexperimental lists were constructed in which the target consonants occurred in filler syllables with a schwa context. For these items, subjects had to detect the target consonants but the responses were not analyzed. These syllables were placed in the antepenultimate syllable ([V, VC, V, V]) in the lists, which made the position of target-bearing syllables in lists unpredictable. In addition to these 45 lists, 45 other lists were constructed as foils which had no instances of the specified target (thus no response was required). Among these 45 foil lists, 30 lists were the same as the 30 experimental lists except that the specified targets did not match the coda consonants in the lists. Each foil list was always presented after its occurrence as an experimental (target-bearing) list. The remaining 15 foil lists contained only vowels in the list ([V, V, V, V]). Finally, 15 practice lists were made.

3. Procedures

The speech materials (V's and VC's) were spoken individually by a male native speaker of Dutch (a speaker from Brabant), and by a male native speaker of American English (a speaker from the Midwest). Both speakers were not linguists, and both were naive as to the purpose of the experiment. The materials were recorded in a sound-treated studio directly onto computer and consequently down-sampled to 22.05 kHz (16 bit precision). All V's and VC's were repeated four times in blocks in a pseudorandomized order, and tokens with deviant prosody (e.g., extreme rising or falling intonation) were excluded. Both speakers released the stops naturally in all instances, though they were not instructed to do so. Randomly selected syllables were then combined to form four-syllable stimulus lists using the Praat speech editor (http://www.praat.org). In the stimulus lists, beginnings of syllables were separated by one second from each other. Two sets of stimuli were constructed which were identical except that, in one set, the release portions of the oral consonants in the VC stimuli were intact, while in the other set the release portions were spliced out (at a zero-crossing at the end of the vowel). The duration of the release portions were measured, as will be discussed in Sec. III B. The intensity of tokens across stimuli was then equalized, so that each stimulus had the same peak intensity but the relative amplitude within a stimulus remained intact.

The task of the subjects was to detect a pre-specified target phoneme irrespective of its position within a spoken
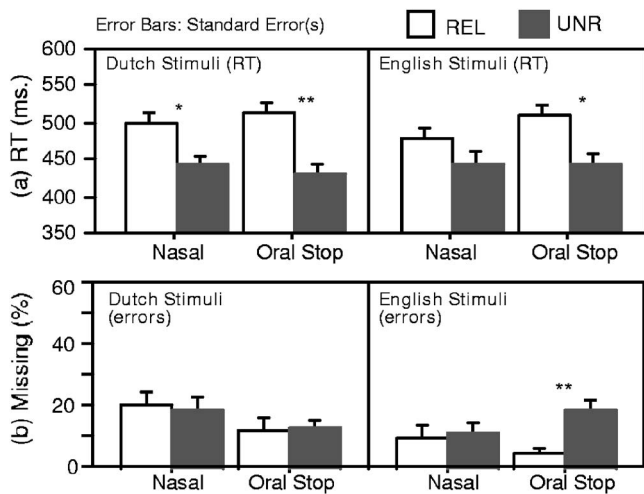
FIG. 1. Korean listeners' mean reaction Times [RTs, (a)] and errors [percent missing responses, (b)] with the spoken stimuli in Dutch (left) vs English (right). REL=released oral stops; UNR=unreleased oral stops. Note the nasal stimuli were identical in the two release conditions. *=$p<0.005$; **=$p<0.001$.

stimulus list. The targets were presented visually in lower-case letters in Roman alphabet for 1 s on a computer screen: **p**, **t**, **k**, **m**, **n**, **ng**, **l**. (/l/ was used in some filler lists.) Note that subjects were sufficiently familiar with the Roman alphabet for targets to be specified in this way. The first syllable of each stimulus list began 300 ms after the target disappeared. A new visual target was presented prior to each list. Subjects were told that they were going to hear some foreign speech (the identity of the language was not revealed) and were instructed to press a button on a response box (with their dominant hand) as fast and as accurately as possible when they detected the targets in the spoken lists.

Participants were tested individually in a quiet room. The computer clock was triggered in synchrony with the onset of the target presentation on the screen and stopped when the response button was pressed. Reaction times (RTs, relative to the offset of the vowel preceding the target phoneme) and errors were recorded. The session lasted approximately 15 min.

4. Analysis

Analyses of Variance (ANOVAs) were performed separately for each language (Dutch/English) with either subjects (F1) or items (F2) as the repeated measure. Both RT and error analyses were carried out (the errors were analyzed in terms of % missing responses). Stop release was a between-subject and within-item factor; manner (Nasal/Oral) was a within-subject and between-item factor. Planned pairwise comparisons between conditions were conducted with separate one-way ANOVAs. Reactions to both oral and nasal targets were analyzed in order to examine the effect of the presence/absence of releases on detection of nasal consonants. Remember that the nasal targets were identical in the released and unreleased conditions.

B. Results

The results are summarized in Fig. 1. Our analyses focus

on the manner of articulation of the target consonants (nasals versus stops) and on whether or not the stops were released. Since our hypotheses did not concern either the place of articulation of the targets or the effects of vowel context, we collapsed over those factors. The patterns we report, however, were generally consistent over both place of articulation and vowel context, as shown in the Appendix.

1. Experiment 1a: Korean natives listening to Dutch

Korean listeners, when presented with Dutch spoken stimuli, showed a significant effect of Release (F1[1,24] =6.969, $p<0.025$; F2[1,28]=28.29, $p<0.001$), such that detection of the targets was faster by 69 ms in the unreleased than in the released condition [mean RTs, 437 vs 506 ms; s.e., 9.5 vs 10.7; see Fig. 1(a)]. Targets were therefore detected faster when they were phonologically viable, even though they were phonetically poorer. Interestingly, however, there was no significant interaction between Release and Manner (F1[1,24]=2.04, $p=0.166$; F2[1,28]<1), suggesting not only that unreleased stops were detected more rapidly than released stops, but also that nasals which were mixed with unreleased stops were processed more rapidly than those mixed with released stops (mean RTs for stops, 431 vs 513 ms; s.e., 14.4 vs 14.9; for nasals, 443 vs 499 ms; s.e., 11.2 vs 12.3). With respect to errors, neither a main effect of Release (F1[1,24]<1; F2[1,28]<1) nor an interaction between Release and Manner (F1[1,24]<1; F2[1,28]<1) was found.

2. Experiment 1b: Korean natives listening to English

In RTs with English stimuli, there was a significant main effect of Release (F1[1,24]=9.71, F2[1,28]=38.72, both p <0.001) and a significant Release × Manner interaction (F1[1,24]=7.86, $p<0.025$; F2[1,28]=4.86, $p<0.05$). As was the case with Dutch stimuli, Korean listeners were significantly faster by 68 ms in detecting (phonologically viable but phonetically poorer) English unreleased stops than released stops (F1[1,24]=4.12, $p<0.05$; F2[1,14]=21.12, p <0.001; mean RTs, 442 vs 510 ms; s.e., 16.4 vs 14.4), as shown in Fig. 1(a) (right). Detection of nasals was also faster in the unreleased than in the released condition by 36 ms, but this effect was significant only by item (F1[1,24]=2.38, p >0.1; F2[1,14]=25.35, $p<0.001$; mean RTs, 445 vs 480 ms; s.e., 11.2 vs 12.3).

In contrast to the RT data, however, detection of targets was more accurate for the (phonetically richer but phonologically nonviable) released stops than for the unreleased stops [see Fig. 1(b)]. There was a main effect of Release (F1[1,24]=9.71, $p=0.005$; F2[1,28]=6.69, $p<0.025$) and a significant Release × Manner interaction (F1[1,24]=7.86, $p=0.01$; F2[1,28]=5.02, $p<0.05$). Planned pairwise comparisons showed that the interaction was due to significantly higher accuracy for the released stops (5% errors; s.e., 1.6) than for the unreleased stops (19% errors; s.e., 2.7; F1[1,24]=22.29, $p<0.001$; F2[1,14]=7.05, $p<0.025$); there was no difference between the nasals in the released and unreleased conditions (F1[1,24]<1; F2[1,14]<1).

T. Cho and J. M. McQueen: Phonological vs phonetics in non-native listening

3. Combined analyses: Experiments 1a and 1b

We also asked whether Korean listeners' phoneme monitoring performance on the stops would be modulated by the language of the stimuli (Dutch versus English). This was a between-subject but within-item factor in two-way (Language $\times$ Release) ANOVAs. In RT, there was again a significant effect of Release (F1[1,48]=11.17, $p<0.005$; F2[1,28]=50.53, $p<0.001$), whereas neither a main effect of Language (F2[1,48]<1; F2[1,28]<1) nor an interaction between Language and Release (F1[1,48]<1; F2[1,28]<1) was found. The error analysis also showed a significant main effect of Release by subject (F[1,48]=9.98, $p<0.005$), but not by item (F2[1,28]=2.73, $p>0.1$) with no main effect of Language (F1[1,48]<1; F2[1,28]<1). There was also a significant interaction between the two factors by subject (F1[1,48]=8.45, $p<0.05$) but not by item (F2[1,28] =2.73, $p>0.1$). Planned pairwise comparisons suggested that this interaction was due to the fact that accuracy in detecting released stops was significantly higher with English (5% errors; s.e., 1.4) than with Dutch syllables (13% errors; s.e., 3.1) (F1[1,24]=5.96, $p<0.025$; F2[1,28]=6.47, $p<0.025$), while there was no language effect on unreleased stops (F1[1,24]=2.81, $p>0.1$; F2[1,28]<1, $p>0.5$; English versus Dutch mean errors: 19% vs 13%; s.e., 2.7 vs 2.0).

C. Discussion

The latency results showed that Korean listeners processed unreleased stops more rapidly than released stops, even though released stops carry additional phonetic cues to phoneme identity. This was true for both Dutch and English stimuli: As far as the RT data are concerned, the results thus appear to support the view that, in non-native listening, sounds which are phonologically viable in the native language are processed more efficiently than sounds which are phonologically illegitimate in the native language.

Interestingly, however, with respect to errors with English stimuli, accuracy in phoneme detection by Korean listeners improved when the targets were released. In other words, released stops (even though they are nonviable in Korean) were more accurately detected than unreleased stops (in spite of the fact that they are viable in Korean). This suggests that when the acoustic-phonetic cues in the release component of the stops were present, errors on the English spoken stimuli decreased, in spite of the illegality of released codas in Korean. This is at least in part in line with the view that the more cues to segment identity there are in the speech signal, the more efficiently the segment is processed (the phonetic-superiority hypothesis). There appears to be a language factor involved, however, since the phonetic-superiority effect in detection accuracy was observed with spoken stimuli in English—which Koreans are familiar with—but not with Dutch—a language unfamiliar to Koreans. This was further confirmed by the combined analysis which showed that the Koreans' detection accuracy was significantly higher for English than for Dutch consonants only in the released condition. We return to this issue in the General Discussion.

III. EXPERIMENT 2

Experiment 1 (the same materials and procedure) was re-run with Dutch listeners. This experiment had two goals. First, it provided a further test of the phonological-superiority hypothesis. If our interpretation of Experiment 1 is correct, such that Koreans detect unreleased stops more rapidly than released stops because unreleased coda stops are viable in Korean, then Dutch listeners should produce the opposite pattern of responses on the same stimuli. They should detect the released stops more easily than the unreleased stops because released coda stops are viable in Dutch. But they may also do so, of course, because the released stops are phonetically richer. Better performance by Dutch listeners on the released stops could thus be due to either phonological or phonetic influences. The critical analysis, therefore, is not that within the Dutch listeners' data itself, but that comparing the Dutch and Korean data (i.e., Experiment 1 versus Experiment 2). In this comparison, phonetic influences are controlled (the stimuli are the same for both groups of listeners), so if the Dutch and Korean listeners were to perform differently, this would support the phonological-superiority hypothesis. We therefore conducted an analysis in which the results of the two experiments were directly compared. Second, Experiment 2 again tested for effects of language familiarity, but now by contrasting performance on native language stimuli (Dutch) with non-native stimuli (English). This contrast provides a measure of whether non–native phoneme monitoring differs from the way the task is performed on native language materials.

A. Method

Seventy-two student volunteers from the Max Planck Institute subject pool were paid for their participation. They were all native speakers of Dutch, with no known hearing disorders. They were divided into four groups of 18, with two groups in Experiment 2a (Dutch stimuli) and two groups in Experiment 2b (English stimuli). In each sub-experiment, one group heard lists with released stops and the other heard lists with unreleased stops. The four sets of stimuli from Experiment 1 were re-used in Experiment 2. All experimental procedures were identical to those in the previous experiment.

B. Results

Ten subjects (five in each sub-experiment) were excluded from the analysis because they made errors so frequently that no scores were recorded for each of those subjects for at least one target consonant.

1. Experiment 2a: Dutch natives listening to Dutch

The RT analyses revealed that Dutch listeners showed the opposite pattern of performance on the Dutch stimuli to that of the Koreans. There was a main effect of Release on RT (F1[1,29]=4.80, $p<0.05$; F2[1,28]=58.55, $p<0.001$), which interacted with Manner (F1[1,29]=19.32, $p<0.001$; F2[1,28]=23.02, $p<0.001$). As shown in Fig. 2(a) (left), the interaction came from the fact that target stops were detected
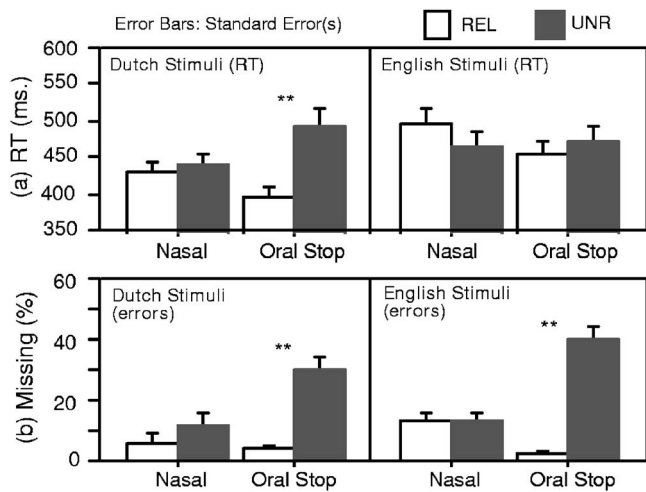
FIG. 2. Dutch listeners' mean Reaction Times [RTs, (a)] and errors [percent missing responses, (b)], with the spoken stimuli in Dutch (left) vs English (right). REL=released oral stops; UNR=unreleased oral stops. Note that the nasal stimuli were identical in the two release conditions. ** = $p < 0.001$.

more rapidly when released than unreleased by 97 ms ($F1[1,29]=13.09$, $p < 0.001$; $F2[1,14]=87.69$, $p < 0.001$; mean RTs: 400 vs 497 ms; s.e., 9.4 vs 17.9), while detection of the nasal targets was not influenced by the release condition—that is, irrespective of whether the nasals were presented in the same list as the released stops or mixed with unreleased stops ($F1[1,29]<1$; $F2[1,14]=4.65$, $p > 0.07$; mean RTs for nasals: 432 vs 443 ms; s.e., 11.9 vs 11.6). Results with respect to the error rates showed a similar pattern (main effect of Release: $F1[1,29]=52.45$, $p < 0.001$; $F2[1,28]=16.71$, $p < 0.001$; Release × Manner interaction, $F1[1,29]=13.94$, $p < 0.001$; $F2[1,28]=6.18$, $p < 0.02$). Detection of stop targets was significantly more accurate when stops were released (4% errors; s.e., 1.3) than when they were unreleased (30% errors; s.e., 3.6; $F1[1,29]=54.89$, $p < 0.001$; $F2[1,14]=13.60$, $p < 0.005$), while there was only a nonsignificant trend for nasal targets (i.e., $F1[1,29]=3.49$, $p > 0.07$; $F2[1,14]=3.11$, $p > 0.09$; mean % errors for nasals: 6% vs 12%; s.e., 1.3 vs 2.9), as can be seen in Figure 2(b) (left).

2. Experiment 2b: Dutch natives listening to English

Unlike when they processed stimuli of their native language, Dutch listeners hearing English stimuli were not faster to respond to released than to unreleased stops [Fig. 2(a), right]. In the RT analysis there was no main effect of Release ($F1[1,29]<1$; $F2[1,28]<1$), but there was a significant Release × Manner interaction ($F1[1,29]=5.689$, $p < 0.025$; $F2[1,28]=7.62$, $p < 0.01$). This interaction was due to the fact that responses were faster for released than for unreleased stops (mean RTs: 459 vs 475 ms; s.e., 12.2 vs 17.8), while this effect reversed for the nasals (mean RTs: 499 vs 471 ms; s.e., 14.2 vs 14.4). But these simple effects were not significant by subjects ($F1[1,29]<1$, for both stops and nasals) or items ($F2[1,14]=3.38$, $p > 0.08$ for stops; $F2[1,14]=4.28$, $p > 0.05$ for nasals).

In errors, however, there was a significant effect of Release ($F1[1,29]=52.91$, $p < 0.001$; $F2[1,28]=23.83$, p

< 0.001) and a significant Release × Manner interaction ($F[1,29]=69.51$, $p < 0.001$; $F2[1,28]=20.59$, $p < 0.001$). The interaction arose because Dutch listeners' detection of English oral stops was far more accurate for released than unreleased stops ($F1[1,29]=116.16$, $p < 0.001$; $F2[1,14]=24.88$, $p < 0.0001$; mean % errors, 2% vs 42%; s.e., 1 vs 4.4) whereas no such effect was found for nasal targets (both F's <1; mean % errors, 13% vs 14%; s.e., 2.4 vs 2.5).

3. Combined analyses: Experiments 2a and 2b

We tested if Dutch listeners' phoneme monitoring performance on the stops would be modulated by the language familiarity factor (Dutch versus English). Similar to Experiment 1, this was a between-subject factor but a within-item factor in two-way (Language × Release) ANOVAs. RT analyses showed no main effect of Language ($F1[1,58]<1$; $F2[1,28]=2.09$, $p > 01$) and a trend toward a Language × Release interaction which was not significant by subject ($F1[1,58]=3.60$, $p=0.06$) but highly significant by item ($F2[1,28]=20.43$, $p < 0.001$). The interaction came from the fact that in processing released (phonological viable) stops, Dutch listeners were faster by 59 ms with native stimuli than with non-native English stimuli ($F1[1,33]=5.75$, $p < 0.025$; $F2[1,28]=10.87$, $p < 0.003$; Dutch versus English mean RTs, 400 vs 459 ms; s.e., 14.6 vs 20.1), while a nonsignificant opposite pattern was found in processing unreleased (phonological unviable) stops (both F's <1; Dutch versus English mean RTs, 497 vs 476 ms; s.e., 24.3 vs 26.9).

Durational analysis of the released stimuli revealed, however, that the duration of the target consonant (the stop closure and the following release noise) was longer for the English than for the Dutch stimuli (closure, 112 vs 102 ms; release noise, 115 vs 90 ms; closure and release combined, 227 vs 192 ms). We therefore asked whether the fact that the Dutch listeners responded more slowly to the English released stops could be attributable to this durational difference. It is possible that listeners' responses could have been slower when the duration of the target consonant was longer, if their decisions tended to be delayed until they had heard all the available phonetic cues. We tested this possibility by examining, for each subject's data, the correlation between consonant duration and RTs. Variation in RTs was not significantly correlated with consonant duration in most cases. For Experiment 2a, only three out of 18 Dutch listeners (with Dutch stimuli) showed a significant positive correlation (the longer the consonant duration, the slower the RT) at $p < 0.05$ (two-tailed), and five listeners showed negative correlation coefficients. For Experiment 2b, none of the Dutch listeners (with English stimuli) showed a significant correlation, and 13 out of 18 listeners showed negative correlation coefficients, which suggests, if anything, that responses were faster when the target consonants were longer. These results indicate that the Dutch listeners' slower responses to the English released stops were not due to the durational difference between the English and Dutch stop releases.

The combined error analysis in Experiment 2 showed a significant main effect of Language ($F1[1,58]=4.11$, $p < 0.05$; $F2[1,28]=38.10$, $p < 0.001$), such that overall accuracy was significantly higher for Dutch than for English

stimuli (Dutch versus English mean errors, 17% vs 22%; s.e., 1.8 vs 1.8). There was, however, a Language × Release interaction by subject (F1[1,58]=7.29, $p<0.01$) but not by item (F2[1,28]=1.49, $p>0.2$). Planned pairwise comparisons showed that this trend was due to the fact that when stops were released, there was no accuracy difference between Dutch and English stimuli (both F's <1; Dutch versus English mean errors, 4% vs 2%; s.e., 1.5 vs 1.0); when stops were unreleased, however, accuracy was significantly higher (by 12%) for Dutch stimuli than for English stimuli by subject (F1[1,25]=5.14, $p<0.04$; mean errors, 30% vs 42%; s.e., 3.6 vs 3.9), but again not by item (F2[1,28]=1.09, $p>0.3$).

4. Combined analyses: Korean and Dutch listeners

Finally, two sets of ANOVAs were run with listener group (Korean versus Dutch) and release condition as independent variables, one with Dutch stimuli (Experiments la and 2a) and one with English stimuli (Experiments 1b and 2b).

In the RT analysis with Dutch stimuli, there was a significant interaction between Listener Group and Release (F1[1,53]=20.59, $p<0.001$; F2[1,28]=98.98, $p<0.001$). The interaction was due to reverse patterns across the two listener groups. In processing Dutch released stops, Dutch listeners detected the released stops more rapidly than Korean listeners did (F1[1,29]=19.15, $p<0.001$; F2[1,28] =43.92, $p<0.001$). The reverse was true with unreleased stops: Korean listeners detected the unreleased stops more rapidly than Dutch listeners did (F1[1,24]=4.78, $p<0.04$; F2[1,28]=6.55, $p<0.02$). In the corresponding error analysis, there was also a significant Listener Group×Release interaction (F1[1,53]=25.31, $p<0.001$; F2[1,28]=7.10, $p<0.02$). This interaction was again due to reverse patterns of performance: the Dutch were more accurate in detecting released stops (F1[1,29]=7.66, $p=0.01$; F2[1,28]=7.28, $p<0.02$); the Koreans were more accurate in detecting unreleased stops, significantly so by subject (F1[1,24]=16.97, $p<0.001$) and marginally so by item (F2[1,28]=3.40, $p<0.08$).

In the analyses with English stimuli, there was a trend towards a Listener Group×Release interaction in the RT analysis by subject (F1[1,53]=3.10, $p<0.08$), but the interaction was highly significant by item (F2[1,28]=22.37, $p<0.001$). This interaction was again due to opposite patterns in the two listener groups. For the released stops, the difference between listener groups was significant only by item (F1[1,28]=2.75, $p>0.1$, F2[1,28]=5.83, $p<0.025$); for the unreleased stops this difference was not significant (F1[1,25]<1; F2[1,28]=3.39, $p<0.08$).

In the corresponding error analysis, however, there was a significant Listener Group×Release interaction (F1[1,53]=27.23, $p<0.001$; F2[1,28]=6.74, $p<0.02$). This interaction did not reflect opposite patterns in the two listener groups, however: the Koreans were more accurate on the unreleased stops than the Dutch (F1[1,25]=23.36, $p<0.001$; F2[1,28]=4.73, $p<0.04$), but there was no accuracy difference between the two groups on the released stops

(F1[1,28]=1.88, $p>0.1$; F2[1,28]=1.33, $p>0.2$). This is probably attributable to a floor effect, given that both listener groups were very accurate.

C. Discussion

Perhaps unsurprisingly, Dutch listeners detected released stops in Dutch syllables faster than in English syllables. Similarly, they detected unreleased stops in the Dutch stimuli more accurately than in the English stimuli. Both of these effects reflect global benefits in processing native over non-native speech. The more specific comparison of listening performance on released and unreleased stops, however, provides evidence on the role of phonological knowledge in phoneme monitoring. When presented with native speech materials, Dutch listeners detected released stop targets more rapidly and more accurately than unreleased stop targets. But when they were presented with English spoken stimuli this effect was attenuated: although it was significant in the error analysis, there was no effect in the RT analysis. A violation of the phonological constraints of the listener's native language thus appears to interfere more severely with processing native than non-native speech.

The overall better performance on released stops can be accounted for in terms of the phonological-superiority hypothesis: Speech processing may be better when the sounds are in harmony with the phonological constraints of the listener's native language than when the sounds are phonologically nonviable. It is also possible, however, that the improved performance in detecting released stops is simply due to the richness of cues present in the signal: The more cues to segment identity in the speech signal, the more efficiently the segment is processed (the phonetic-superiority hypothesis). When a stop is released, its identity is cued not only by the formant transition in the vowel but also by the spectral characteristics during the burst, whereas an unreleased stop is perceptually weaker because it is cued only by the formant transition information.

One might argue that the attenuation of the released/ unreleased difference in non-native listening (i.e., the Dutch listening to English) offers support for the phonological-superiority hypothesis. If the difference is due to phonological knowledge about the native language, one might expect a weaker effect when the input consists of non-native syllables. But it remains possible that the phonetic cues were stronger in the Dutch than in the American English stimuli. The critical comparisons are therefore the cross-experiment contrasts (i.e., on the Korean versus Dutch listener groups), where phonetic differences, should they exist, are held constant. These comparisons indeed showed that there were significantly different patterns in the two listener groups on both sets of stimuli. This suggests that phonological viability in the native language plays a stronger role in non-native listening than the availability and/or relative richness of acoustic-phonetic cues.

IV. GENERAL DISCUSSION

On the basis of the results of two experiments, we draw three conclusions. First, in line with previous research, per-

ception of non-native and native speech is influenced by the phonological system of the listener's native language. Second, although the richness of acoustic phonetic information also plays a role in consonant identification, this influence in non-native perception can be weaker than that due to native language phonological knowledge. Third, perception of non-native speech is also influenced by the listener's familiarity with the phonetic detail in the non-native language. We discuss each of these points in turn.

In line with the phonological-superiority hypothesis, phoneme recognition is improved in a speech environment which is phonologically viable in the listener's native language: Korean listeners detected unreleased coda stops faster than released coda stops, but Dutch listeners, responding to the same stimuli, detected the released stops more rapidly and/or more accurately. This pattern was confirmed by our cross-experiment comparisons, which showed that, for the unreleased Dutch stops, the Koreans were faster and more accurate than the Dutch, whereas for the released Dutch stops, the Dutch were faster and more accurate than the Koreans. The same pattern, though statistically weaker, was also observed for the English stimuli; we discuss the reasons for the weakening of the effect in the following. Nevertheless, there was still an influence of native phonology in the error rates on the English unreleased stops: Korean listeners missed significantly fewer target phonemes in this condition than the Dutch listeners. Finally, the results indicate that this phonological effect was stronger in native listening (Dutch listeners, Dutch VC's; significant effects in both speed and accuracy) than in non-native listening (Dutch listeners, English VC's; significant effects only in accuracy).

We suggest that listeners' expectations about the acoustic-phonetic form of a stop consonant in coda position, as determined by the phonological sound patterns of the listeners' native language and therefore a lifetime's exposure to those sound patterns, influence ease of phoneme recognition in native and, more strikingly, in non-native listening. Previous research has shown that various aspects of the phonological structure of one's native language influence non-native listening [e.g., phoneme inventory, Costa *et al.* (1998), Flege and Hillenbrand (1986); place assimilation rules, Weber (2001); phonotactic restrictions, Broersma (2005), Weber and Cutler (2006); and rhythmic structure, Cutler *et al.* (1986), Cutler and Otake (1994), Otake *et al.* (1993)]. The present findings add to this body of evidence by showing that native-language experience can determine ease of phoneme identification even in the extreme case where non-native target consonants which are phonologically expected in the listener's native language are in fact phonetically poorer (as when the Korean listeners detected unreleased stops more readily than released stops).

This, then, is our second conclusion: phonological expectations can outweigh phonetic information. It is undeniably true that native and non-native listening depends on the uptake of acoustic-phonetic cues present in the signal. In general, one would therefore assume that speech sounds which are more richly specified (e.g., released coda stops cued by VC formant transition information and by the information in the release burst) would be easier to recognize than

sounds which are less well specified (e.g., unreleased coda stops cued only by the VC formant transitions). The present results suggest, however, that at least under some circumstances, poorer acoustic-phonetic information can lead to better phoneme recognition in non-native listening.

The relevant phonological fact that has been the focus of the present study is that Korean stops are obligatorily unreleased in coda position in words spoken in isolation, as compared to Dutch stops, which are generally released in that environment. There is, however, an additional phonological difference concerning coda position in Korean and Dutch which might be relevant. Korean stops are phonologically divided into the categories lenis, fortis and aspirated, and these manners of articulation are cued primarily in the post-release part of the acoustic signal (Cho *et al.*, 2002; Kim *et al.*, 2002). This three-way manner contrast is neutralized completely in coda position (Kim and Jongman, 1996). Furthermore, the three fricatives in Korean (/s,s*,h/) become unreleased alveolar stops in coda position (Kim-Renaud, 1976; Martin, 1992). The situation is very different in Dutch codas. Although there is word-final devoicing in Dutch, this process is incomplete: There are fine-grained phonetic differences between underlyingly voiced and voiceless coda stops (Warner *et al.*, 2004; Ernestus and Baayen, in press). Because of this incomplete neutralization process, five stops can appear in Dutch singeleton codas. Fricatives can also occur in the coda in Dutch. The possible phonetic forms that can occur in coda position are thus more limited in Korean than in Dutch. The functional load in processing consonants in the coda may therefore be lower for Korean than for Dutch listeners. It is possible that this difference may influence perception of coda consonants in the two languages. Our findings do not bear directly on this issue, however, because phonologically viable phonetic forms in the coda position are unreleased in Korean but released in Dutch, and it is this which we manipulated, not the size of the set of coda consonants. A pure test of this set-size hypothesis would require a comparison of two languages with the same type of stops in the coda (either released or unreleased), so that the only relevant factor to consider would be the different number of allowable phonetic forms in the coda stemming from language-specific neutralization processes.

Although poorer acoustic-phonetic information led to faster phoneme recognition in non-native listening by the Koreans, there was some evidence that acoustic-phonetic cues can also be exploited by listeners in non-native speech perception, even if those cues are not the ones used in processing the listener's native language. Koreans were faster on unreleased than on released stops in both Dutch and English. But, for the English stimuli alone, the Koreans were more accurate on the released than on the unreleased stops. We suggest that this interaction between the application of phonological constraints and the use of acoustic-phonetic cues in non-native perception depends on language familiarity. Dutch is a truly exotic language to Korean university students, whereas English is, for them, the most familiar non-native language. They are likely to have learned English since they were 12 years old but never to have heard Dutch before. In the course of second-language acquisition, Korean

T. Cho and J. M. McQueen: Phonological vs phonetics in non-native listening

students might have built up some familiarity with the acoustic-phonetic characteristics of the sounds of English. If so, then even though the violation of the phonological constraints of Korean in English released stops led to slow detection times, familiarity with the form of such stops may have allowed Korean learners of English to exploit the richer set of acoustic-phonetic cues available in the released stops to improve detection accuracy.[1]

On this view, Korean listeners are able to extract efficiently information about stop identity in the formant transitions of the preceding vowel. When there is no release (in both the Dutch and English stimuli), they are therefore able to recognize target phonemes quickly, and quite accurately. When a release follows, however, their phoneme monitoring performance is disrupted, resulting in slower responses to both the Dutch and English released stimuli. Familiarity with English can nevertheless allow Koreans to be very accurate on the English released stops. They have no experience with Dutch and thus the impact of the release-induced disruption is greater, that is, the additional cues in the Dutch released stops are not fully exploited. But the additional information in the English releases, though unable to undo the disruption in processing speed caused by the releases' phonological nonviability, leads to very accurate performance. Thus, if non-native speech contains cues that are not present in the equivalent position in native speech, and if non-native listeners are sufficiently familiar with that language, then those listeners can use those richer cues in phoneme identification.

This use of non-native cues contrasts with the findings of Broersma (2005). She showed that Dutch listeners who were very familiar with English (at least as familiar as the present Koreans were with English) did not process vowel-duration cues to the English coda voicing distinction, unlike a group of native English listeners. The Dutch listeners nevertheless categorized non-native English coda fricatives as voiced or voiceless at least as well as the native English listeners. Broersma suggests that this may have been because the Dutch listeners have learned that vowel duration is not a very informative cue to the voicing distinction in Dutch, and thus do not rely on this cue in non-native listening either. The present results suggest however that it is not necessarily the case that phonetic information that is not valuable in native language perception (or that is even not present in the native language) is ignored when the listener hears a foreign language. One reason for the difference between studies may be due to a difference in cue exposure. The Dutch listeners in the Broersma study have been exposed to differences in vowel duration in their native language, as a cue to vowel quantity (i.e., the phonemic contrast between long and short vowels; cf. Nooteboom, 1972; Nooteboom and Doodeman, 1980) and to other cues to the obstruent voicing contrast. Dutch listeners may thus have learned that vowel duration is

a weak cue to obstruent voicing. But the Korean listeners in the present study are never exposed to release bursts in coda position in their native language and therefore cannot learn, during native language exposure, to ignore this information in stop identification or to use it for the recognition of some other sound(s). When Koreans therefore have sufficient exposure to release bursts in non-native listening, they can use the additional information contained in those bursts to improve target recognition accuracy.

The present study has pitted native-language phonological viability against the relative richness of acoustic-phonetic cues in non–native listening. Phonological viability seems to have the upper hand: Phoneme recognition is in general better when the phonetic specification of those phonemes is in keeping with the listener's native phonology, even when that phonetic specification is, in physical terms, poorer. This dominance of phonological knowledge is consistent with theories of non-native speech perception, such as Flege's Speech Learning Model (SLM, e.g., Flege, 1991, 1995) and Best's Perceptual Assimilation Model (PAM, e.g., Best, 1994, 1995, Best *et al.*, 2001). Both models predict that nonnative sounds are perceived through some form of phonological filter of the listeners' native language that takes into account the nonnative sounds' phonetic (dis)similarities to native phonetic categories. Previous studies supporting these models, however, have generally been based on cross-linguistic phonological differences concerning the phoneme inventories of the languages in question. The findings of the present study therefore show that the phonological filter created through exposure to a native language is determined not only by the phoneme inventory of that language but also by sound patterns that are driven by phonological (allophonic) rules. As we have just argued, however, the relative richness of phonetic cues can also play a role in non-native listening, at least in a non-native language with which the listeners are familiar. Models of non-native listening must therefore not only take native-language phonological processes into account (both phonemic and rule-driven processes), but also the absolute amount of phonetic information in the nonnative speech stream.

ACKNOWLEDGMENTS

APPENDIX

TABLE I. Mean reaction times (ms) by place of articulation (/p,t,k/) and vowel context (/a,e,i,o,u/).

Listeners	Stimuli	Vowel context	/p/ REL	/p/ UNR	/t/ REL	/t/ UNR	/k/ REL	/k/ UNR
Korean listeners (Experiment 1)	Dutch	/a/	462	394	556	516	480	473
		/e/	479	390	574	498	557	392
		/i/	451	343	546	453	419	347
		/o/	550	395	524	568	478	373
		/u/	472	443	508	432	629	537
	English	/a/	423	380	444	426	478	370
		/e/	501	436	513	407	499	492
		/i/	557	446	533	601	444	430
		/o/	473	391	608	490	544	383
		/u/	528	423	485	389	616	551
Dutch listeners (Experiment 2)	Dutch	/a/	431	490	419	568	363	518
		/e/	371	427	433	495	398	524
		/i/	386	475	448	476	407	509
		/o/	360	421	394	532	368	428
		/u/	367	457	379	489	473	580
	English	/a/	420	411	408	442	358	468
		/e/	417	445	478	486	433	412
		/i/	477	479	510	560	469	501
		/o/	402	409	523	475	473	501
		/u/	480	520	426	541	602	560

NOTE: REL=released stops; UNR=unreleased stops.

TABLE II. Percent errors (missing responses) by place of articulation (/p,t,k/) and vowel context (/a,e,i,o,u/).

Listeners	Stimuli	Vowel context	/p/ REL	/p/ UNR	/t/ REL	/t/ UNR	/k/ REL	/k/ UNR
Korean listeners (Experiment 1)	Dutch	/a/	8	15	8	38	31	0
		/e/	8	0	23	0	8	0
		/i/	8	0	8	8	8	8
		/o/	0	8	38	8	8	8
		/u/	8	0	15	8	15	92
	English	/a/	0	8	0	8	0	0
		/e/	15	15	0	15	8	8
		/i/	0	0	8	38	0	0
		/o/	0	8	8	8	8	69
		/u/	0	15	0	8	23	85
Dutch listeners (Experiment 2)	Dutch	/a/	17	15	0	38	0	15
		/e/	0	15	22	69	0	8
		/i/	0	0	6	46	0	46
		/o/	0	0	0	62	11	8
		/u/	6	23	0	15	0	92
	English	/a/	0	7	6	29	6	21
		/e/	0	7	0	43	6	14
		/i/	0	21	0	93	6	71
		/o/	0	14	0	79	6	64
		/u/	0	21	0	43	6	93

NOTE: REL=released stops; UNR=unreleased stops.

[1]This language familiarity effect is further supported by an analysis of detection accuracy on the nasals. (We thank an anonymous reviewer for bringing the detection accuracy data on nasals to our attention.) As can be seen in Fig. 1, Korean listeners were more accurate in detecting (for them more familiar) English nasals than (for them unfamiliar) Dutch nasals, especially in the released condition (10% vs 21% errors; $F1_{(1,24)}=7.08$, $p=0.014$; $F2_{(1,14)}=1.93$, $p>0.1$) but also in the unreleased condition (11% vs 20% errors; $F1_{(1,24)}=3.49$, $p=0.174$; $F2_{(1,14)}=1.31$, $p>0.1$). A similar observation can be made from the Dutch listeners' responses to the nasals (Fig. 2). Dutch listeners were more accurate in detecting Dutch nasals than less familiar (non-native) English nasals, but only in the released condition (6% vs 13% errors; $F1_{(1,29)}=6.38$, $p=0.016$; $F2_{(1,14)}=4.45$, $p=0.053$).

Best, C. T., McRoberts, G. W., and Goodell, E. (**2001**). "American listeners' perception of nonnative consonant contrasts varying in perceptual assimilation to English phonology," J. Acoust. Soc. Am. **109**, 775–794.

Best, C. T. (**1995**). "A direct realist view of cross-language speech perception," in *Speech Perception and Linguistic Experience*, edited by W. Strange (York, Timonium, ND), pp. 171–204.

Best, C. T. (**1994**). "The emergence of language-specific phonemic influences in infant speech perception," in *Development of speech perception: The transition from speech sounds to spoken words*, edited by J. Goodman and J. C. Nusbaum (MIT Press, Cambridge, MA), pp. 167–224.

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

Best, C. T., McRoberts, G. W., and Sithole, N. M. (**1988**). "Examination of perceptual reorganization of non-native speech contrasts: Zulu click discrimination by English-speaking adults and infants," J. Exp. Psychol. Hum. Percept. Perform. **14**, 345–360.

Bradlow, A. R., Pisoni, D. B., Arkahane-Yamada, R., and Tohkura, Y. (**1997**). "Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production," J. Acoust. Soc. Am. **101**, 2299–2310.

Broersma, M. (**2005**). "Perception of familiar contrasts in unfamiliar positions," J. Acoust. Soc. Am. **117**, 3890–3901.

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

Cho, T., and McQueen, J. M. (**2004**). "Phonotactics vs. phonetic cues in native and non-native listening: Dutch and Korean listeners' perception of Dutch and English," in *Proceedings of ICSLP2004 (International Conference on Spoken Language Processing), Jeju Island, Korea* (Sunjin, Seoul), pp. 1301–1304.

Connine, C. M., and Titone, D. (**1996**). "Phoneme monitoring," Lang. Cognit. Processes **11**, 635–645.

Costa, A., Cutler, A., and Sebastián-Gallés, N. (**1998**). "Effects of phoneme repertoire on phoneme decision," Percept. Psychophys. **60**, 1022–1021.

Cutler, A., and Broersma, M. (**2005**). "Phonetic precision in listening," in *A Figure of Speech*, edited by W. Hardcastle and J. Beck (Erlbaum, Mahwah, NJ), pp. 63–91.

Cutler, A., Mehler, J., Norris, D., and Seguí, J. (**1986**). "The syllable's differing role in the segmentation of French and English," J. Mem. Lang. **25**, 385–400.

Cutler, A., and Otake, T. (**1994**). "Mora or phoneme? Further evidence for language-specific listening," J. Mem. Lang. **33**, 824–844.

Denes, P. (**1955**). "Effect of duration on the perception of voicing," J. Acoust. Soc. Am. **27**, 761–764.

Ernestus, M., and H. Baayen. (**in press**). "The functionality of incomplete neutralization in Dutch. The case of past-tense formation," in *Laboratory phonology 8: Varieties of Phonological Competence*, edited by L. M. Goldstein, D. H. Whalen, and C. T. Best (Mouton de Gruyter, Berlin).

Flege, J. E. (**1995**). "Second language speech learning: Theory, findings, and problems," in *Speech Perception and Linguistic Experience*, edited by W. Strange (York, Timonium, ND), pp. 133–172.

Flege, J. E. (**1991**). "Orthographic evidence for the perceptual identification of vowels in Spanish and English," Q. J. Exp. Psychol. **43**, 701–731.

Flege, J. E., and Hillenbrand, J. (**1986**). "Differential use of temporal cues to the /s/-/z/ contrast by native and non-native speakers of English," J. Acoust. Soc. Am. **79**, 508–517.

Flemming, E. (**2005**). "Speech perception and phonological contrast," in *The Handbook of Speech Perception*, edited by D. Pisoni and R. Remez (Blackwell), pp. 156–181.

Hallé, P. A., Best, C. T., and Levitt, A. (**1999**). "Phonetic vs. phonological

influences on French listeners' perception of American English approximants," J. Phonetics **27**, 281–306.

Henderson, J. B., and Repp, B. H. (**1982**). "Is a stop consonant released when followed by another stop consonant?," Phonetica **39**, 71–82.

Jusczyk, P. W. (**1997**). *The Discovery of Spoken Language* (MIT, Cambridge, MA).

Kim, H., and Jongman, A. (**1996**). "Acoustic and perceptual evidence for complete neutralization of manner of articulation in Korean," J. Phonetics **24**, 295–312.

Kim, M.-Y., Beddor, P. S., and Horroks, J. (**2002**). "The contribution of consonantal and vocalic information to the perception of Korean initial stops," J. Phonetics **30**, 77–100.

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

Kochetov, A. (**2001**). "Production, perception, and emergent phonotactic patterns: A case of contrastive palatalization," Ph.D. dissertation, University of Toronto.

Kuhl, P. K., and Iverson, P. (**1995**). "Linguistic experience and the perceptual magnet effect," in *Speech Perception and Linguistic Experience*, edited by W. Strange (York, Timonium, ND), pp. 12l–154.

Lively, S. E., Logan, J. S., and Pisoni, D. B. (**1993**). "Training Japanese listeners to identify English /r/ and /l/: II. The role of phonetic environment and talker variability in learning new perceptual categories," J. Acoust. Soc. Am. **94**, 1242–1255.

Lively, S. E., Logan, J. S., Pisoni, D. B., Yamada, R. A., Tohkura, Y., and Yamada, T. (**1994**). "Training Japanese listeners to identify English /r/ and /l/: III. Long-term retention of new phonetic categories," J. Acoust. Soc. Am. **96**, 2076–2087.

Logan, J. S., Lively, S. E., and Pisoni, D. B. (**1991**). "Training Japanese listeners to identify English /r/ and /l/: A first report," J. Acoust. Soc. Am. **89**, 874–866.

Martin, S. E. (**1992**). *A Reference Grammar of Korean* (Tuttle, Rutland, VT).

Mattingly, I. G. (**1981**). "Phonetic representation and speech synthesis by rule," in *Proceedings of the 13th International Congress of Phonetic Sciences*, edited by T. Myers, J. Laver, and J. Anderson (Royal Institute of Technology and Stockholm University, Stockholm), pp. 574–577.

Nooteboom, S. G. (**1972**). "Production and perception of vowel duration," Vol. 4, Ph.D. dissertation. Rijksuniversiteit Utrecht.

Nooteboom, S. G., and Doodeman, G. J. N. (**1980**). "Production and perception of vowel length in spoken sentences," J. Acoust. Soc. Am. **67**, 267–287.

Otake, T., Hatano, G., Cutler, A., and Mehler, J. (**1993**). "Mora or syllable? Speech segmentation in Japanese," J. Mem. Lang. **32**, 358–378.

Otake, T., Yoneyama, K., Cutler, A., and van der Lugt, A. (**1996**). "The representation of Japanese moraic nasals," J. Acoust. Soc. Am. **100**, 3831–3842.

Polka, L., Colantonio, C., and Sundara, M. (**2001**). "A cross-language comparison of /d/-/ð/ perception: Evidence for a new developmental pattern," J. Acoust. Soc. Am. **109**, 2190–2220.

Raphael, L. J. (**1972**). "Preceding vowel duration as a cue to the perception of the voicing characteristics of word-final consonants in American English," J. Acoust. Soc. Am. **51**, 1296–1303.

Silverman, D. (**1995**). "Phasing and Recoverability", Ph.D. dissertation, University of California, Los Angeles.

Smits, R., Ten Bosch, L., and Collier, R. (**1996**). "Evaluation of various sets of acoustic cues for the perception of prevocalic stop consonants. I. Perception experiment," J. Acoust. Soc. Am. **100**, 3852–3864.

Steriade, D. (**1997**). "Phonetics in phonology. The case of laryngeal neutralization," University of California, Los Angeles (Unpublished) (http://web.mit.edu/linguistics/www/bibliography/steriade.html).

Stevens, K. N., and Blumstein, S. E. (**1978**). "Invariant cues for place of articulation in stop consonants," J. Acoust. Soc. Am. **64**, 1358–1368.

Warner, N., Jongman, A., Sereno, J., and Kemps, R. (**2004**). "Incomplete neutralization and other sub-phonetic durational differences in production and perception: Evidence from Dutch," J. Phonetics **32**, 251–276.

Watson, I. (**1983**). "Cues to the voicing contrast: A survey," Cambridge Papers in Phonetics and Experimental Linguistics **2**, 1–34.

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

Weber, A., and Cutler, A. (**2006**). "First-language phonotactics in second-language listening," J. Acoust. Soc. Am. **119**, 597–607.

Werker, J. F., and Lalonde, C. E. (**1988**). "Cross-language speech percep-

tion: Initial capabilities and developmental change," Dev. Psychol. **24**, 672–683.

Werker, J. F., Gilbert, J. H. V., Humphrey, K., and Tees, R. C. (**1981**). "Developmental aspects of cross-language speech perception," Child Dev. **52**, 349–355.

Wright, R. (**2004**). "A review of perceptual cues and cue robustness," in *Phonetically Based Phonology*, edited by B. Hayes, R. Kirchner, and D. Steriade (Cambridge University Press, Cambridge), pp. 34–57.

Zsiga, E. (**2003**). "Articulatory timing in a second language. Evidence from Russian and English," Stud. Second Lang. Acquis. **25**, 399–432.