



Phonotactics vs. phonetic cues in native and non-native listening: Dutch and Korean listeners' perception of Dutch and English

Taehong Cho & James M. McQueen

Max Planck Institute for Psycholinguistics
Nijmegen, The Netherlands
Taehong.Cho@mpi.nl & James.McQueen@mpi.nl

Abstract

We investigated how listeners of two unrelated languages, Dutch and Korean, process phonotactically legitimate and illegitimate sounds spoken in Dutch and American English. To Dutch listeners, *unreleased* word-final stops are phonotactically illegal because word-final stops in Dutch are generally released in isolation, but to Korean listeners, *released* final stops are illegal because word-final stops are never released in Korean. Two phoneme monitoring experiments showed a phonotactic effect: Dutch listeners detected released stops more rapidly than unreleased stops whereas the reverse was true for Korean listeners. Korean listeners with English stimuli detected released stops more accurately than unreleased stops, however, suggesting that acoustic-phonetic cues associated with released stops improve detection accuracy. We propose that in non-native speech perception, phonotactic legitimacy in the native language speeds up phoneme recognition, the richness of acoustic-phonetic cues improves listening accuracy, and familiarity with the non-native language modulates the relative influence of these two factors.

1. Introduction

In the first months of life, infants are able to discriminate acoustic-phonetic differences irrespective of whether the phonetic differences are phonologically contrastive in the ambient language or not (e.g., [1]). However, by the time they are able to say a few words, their sensitivity to speech sound contrasts is refined according to the phonological properties of the native language (e.g., [2], see [3] for a review), such that sensitivity to phonetic differences which are phonologically non-contrastive in the native language becomes attenuated relative to sensitivity to contrastive phonetic differences. Eventually, perception of speech sounds by adults is highly biased or modulated by their experience with the phonological systems of their native language (e.g., phoneme repertoire, phonotactics, allophonic distribution, morpho-phonemic alternations, see, e.g., [4, 5, 6, 7, 8]). Adult native Japanese listeners, for example, generally perceive both contrastive English sounds /r/ and /l/ as the flap /r/, which is the sound in their native language that is phonetically closest to the two sounds ([4]). What has emerged from these studies on native vs. non-native listening then is that language experience alters the mechanisms which underlie speech perception, and that one of the driving forces behind these alterations is the phonological system of the listener's native language.

In the present study, we further explore this issue by investigating how listeners of two unrelated languages, Dutch

and Korean, process sounds which are phonologically legitimate or illegitimate in their native language. In Dutch, stops in the word-final coda are generally released when produced in isolation, whereas in Korean, stops are never released in the coda. The question is then whether and how such a phonotactic difference between the two languages influences speech perception. If native phonotactics play an important role in non-native perception, speech sounds with a violation of such phonotactics will be processed less efficiently than those which respect the native constraints. 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. Thus, from the perspective of the relative richness of acoustic-phonetic cues, one might expect that released stops are processed more efficiently independently of the phonotactics of the listener's native language. To test these hypotheses, two phoneme monitoring experiments were conducted, one with Dutch and one with Korean listeners.

2. Experiment 1

We used the phoneme monitoring paradigm to test if the phonotactic constraint that Dutch word-final stops are usually released influences listening performance. Dutch listeners were presented with Dutch spoken stimuli in the first version of the experiment (Exp. 1a) and with English spoken stimuli in the second version of the experiment (Exp. 1b). Two questions to be addressed were (1) how the speed and accuracy of phoneme monitoring varies as a function of the phonotactic constraint—i.e., whether the final stops are released or unreleased, and (2) whether the phonotactic influence on phoneme monitoring, if any, is further constrained by language experience (native vs. non-native speech stimuli, i.e., Dutch vs. English).

2.1. Method

Materials. Two sets of materials were constructed in each language, one consisting of VC items with the release intact (the released condition), and one with the release spliced out (the unreleased condition). Within each set of VCs, 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 x 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]. The non-target-bearing syllables in the lists consisted

of Vs only. In addition to these 30 lists, 15 non-experimental 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, fifteen practice lists were made.

Procedure and Subjects. The speech materials (Vs and VCs) were spoken individually by a male native speaker of Dutch, and by a male native speaker of American English. They were recorded in a sound-treated studio directly onto computer and consequently down-sampled to 22.05 kHz (16 bit precision). All Vs and VCs were repeated four times in blocks in a pseudo-randomized order, and tokens with deviant prosody (e.g., extreme rising or falling intonation) were excluded. 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 a speech editor. 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).

Seventy-two Dutch student volunteers in Nijmegen were paid to participate. They were divided into four groups of 18, with two groups in Exp. 1a with Dutch stimuli and two in Exp. 1b with English stimuli. In each experiment, one group heard the released lists and the other the unreleased lists.

The task of subjects was to detect a pre-specified target phoneme irrespective of its position within a spoken stimulus list. The targets were presented visually in lowercase letters for one second on a computer screen: **t, p, k, m, n, ng, l.** (/l/ was used in some filler lists.) 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 instructed to press a button as fast and as accurately as possible when they detected the targets in the spoken lists. 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. Response latencies relative to the onset of the target phoneme and errors were recorded. The session lasted approximately 15 minutes. Ten subjects 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.

To test the effect of the presence or absence of the release on phoneme detection, repeated measures ANOVAs were performed separately for each spoken language (Dutch in Exp. 1a, English in Exp. 1b) with Release (REL/UNR) as a between-subject factor and Manner (Nasal/Oral) as a within-subject factor. Reactions to both oral and nasal targets were analyzed in order to examine the effect of Release on

detection of oral stops in comparison with that of nasals, which were identical in the released and unreleased conditions.

2.2. Results

Exp. 1a: Dutch natives listening to Dutch. There was a main effect of Release on RT ($F[1,29]=4.801, p<0.05$), which interacted with Manner ($F[1,29]=19.315, p<0.0001$). As shown in Figure 1a (left), the interaction came from the fact that target stops were detected more rapidly when released than unreleased ($p<0.001$), while detection of the nasal targets was not influenced by the release condition—i.e., irrespective of whether nasal targets were presented in the same group with released stops or with unreleased stops ($p=0.26$). Results with respect to the error rates (percent missing responses) showed a similar pattern (main effect of Release: $F[1,29]=52.45, p<0.0001$; Release x Manner interaction, $F[1,29]=13.94, p<0.001$). Detection of stop targets was significantly more accurate when stops were released than unreleased ($p<0.0001$), while there was only an insignificant trend for nasal targets, as can be seen in Figure 1b (left).

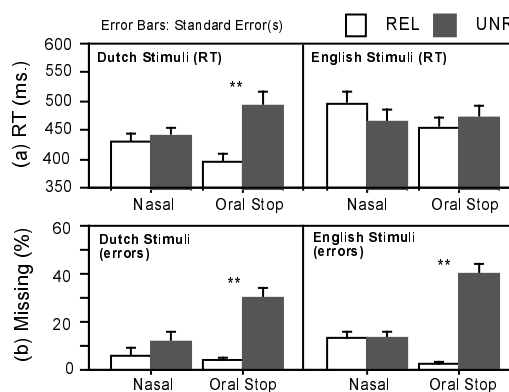


Figure 1. Dutch listeners' mean Reaction Times (RTs, a) and errors (percent missing responses, b), with the spoken stimuli in Dutch (left) vs. English (right). REL = released oral stops; UNR = unreleased oral stops.

Exp. 1b: Dutch natives listening to English. Unlike when they processed Dutch stimuli, Dutch listeners hearing English stimuli were not faster to respond to released than to unreleased stops (Figure 1a, right). In the RT analysis there was no main effect of Release ($F[1,29]=0.035, p=0.852$), but there was a significant Release x Manner interaction ($F[1,29]=5.689, p<0.025$). This interaction was due to the fact that responses were faster for released stops than for unreleased stops, while this effect reversed for the nasal consonants. Neither of these effects was significant, however ($p>0.05$). In errors, on the other hand, there was a significant effect of Release ($F[1,29]=52.911, p<0.0001$) and a significant Release x Manner interaction ($F[1,29]=69.511, p<0.0001$). The interaction arose because Dutch listeners' detection of English oral stops was far more accurate for released than unreleased stops ($p<0.0001$) whereas no such effect was found for nasal targets ($p>0.05$).

2.3. Discussion

The results showed that when presented with native speech materials, Dutch listeners detected the released stop targets

more rapidly and more accurately relative to the unreleased stop targets. This suggests that spoken phonemes are processed more efficiently when the sounds are phonotactically legitimate in the listener's native language (released) than when they are illegal (unreleased). In the second version of the experiment with English spoken stimuli, the phonotactic effect was attenuated: There was no effect of phonotactics in the RT analysis, but it was significant in the error analysis. It appears that a violation of the native language's phonotactics interferes with processing of the native language more severely than with non-native language processing.

The overall better monitoring performance with released stops (vs. unreleased stops) can be accounted for in terms of the hypothesis that speech processing is better when the sounds are in harmony with the phonotactics of the listener's native language than when the sounds are phonotactically illegal. 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. 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.

In Experiment 2, we explored these two possibilities by examining the responses of native Korean listeners to the same speech materials used in Experiment 1. Korean is a language in which stops are never released in the coda position. Therefore, released stops, which were in harmony with Dutch phonotactics, are in disharmony with the phonotactics of Korean. If phonotactics play a stronger role in speech perception than the acoustic-phonetic cues in the signal, Korean listeners should demonstrate the pattern opposite to what Dutch listeners showed in Experiment 1: *unreleased* stops should be detected more rapidly and accurately than released stops. On the other hand, if acoustic-phonetic cues weigh more heavily than phonotactics, Korean listeners should detect released stops more efficiently than unreleased stops. We also examined the influence of non-native language experience on non-native listening: Korean listeners are much less familiar with Dutch than with English.

3. Experiment 2

3.1. Method

The same four sets of spoken stimuli in Exp. 1 were used in Exp. 2. Fifty-two Korean student volunteers at Korea University in Seoul were paid to take part in the experiment. Again, they were divided into four groups of 13 according to Release (Released vs. Unreleased) and Language (Dutch in Exp. 2a and English in Exp. 2b). All other experimental procedures were identical to those in Exp. 1. Note that the subjects were sufficiently familiar with the Roman alphabet for targets to be specified in the same way as in Exp. 1.

3.2. Results

Exp. 2a: Korean natives listening to Dutch. Korean listeners, when presented with Dutch spoken stimuli, showed the exact opposite performance to that of the Dutch listeners

in Exp. 1. There was a significant effect of Release ($F[1,24]=6.969, p<0.025$), such that detection of the targets was far faster in the *unreleased* condition than in the released condition, as shown in Figure 2a (left). Interestingly, however, there was no significant interaction between Release and Manner ($F[1,24]=2.04, p=0.166$), suggesting that, in addition, nasals which were mixed with released stops were processed more rapidly than those mixed with unreleased stops. (Recall that lists with nasal targets were identical between the released and unreleased conditions.) With respect to errors, neither a main effect of Release ($F[1,24]=0.007, p=0.934$) nor an interaction between Release and Manner ($F[1,24]=0.083, p=0.776$) was found.

Exp. 2b: Korean natives listening to English. In RTs with English stimuli, there was a significant main effect of Release ($F[1,24]=9.712, p<0.001$) and a significant Release x Manner interaction ($F[1,24]=7.86, p<0.025$). As was the case with Dutch stimuli, Korean listeners were significantly faster in detecting English unreleased stops than released stops ($p<0.05$), as shown in Figure 2a (right), but this time detection of nasals was not influenced by the release condition ($p>0.05$).

In contrast to the RT data, however, detection of targets was far more accurate for the (illegal) released than for (legal) unreleased condition, as shown in Figure 2b (right). There was a main effect of Release ($F[1,24]=9.712, p=0.005$) and a significant Release x Manner interaction ($F[1,24]=7.86, p=0.01$). The interaction was due to significantly higher accuracy for the released stops (vs. the unreleased stops), but not for the nasals in the released versus the unreleased conditions.

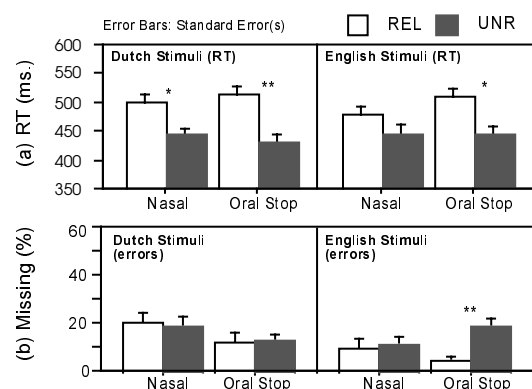


Figure 2. Korean listeners' mean Reaction Times (RTs, a) and errors (percent missing responses, b) with the spoken stimuli in Dutch (left) vs. English (right). REL = released oral stops; UNR = unreleased oral stops.

3.3. Discussion

In Experiment 2, we examined how Korean listeners monitored for released vs. unreleased stops spoken in two non-native languages, Dutch and English. The latency results showed that, unlike Dutch listeners, Korean listeners processed unreleased stops more rapidly than released stops, even if released stops carry additional phonetic cues to phoneme identity. This was true for both Dutch and English stimuli (in Exp. 2a and Exp. 2b). As far as the RT data are concerned, the results then appear to support the view that, in non-native listening, sounds which are phonotactically

legitimate in the native language are processed more efficiently than sounds which are phonotactically illegal in the native language.

Interestingly, however, with respect to errors with English stimuli, the accuracy in phoneme detection by Korean listeners declined when the targets were unreleased. In other words, although Korean listeners processed English unreleased stops more rapidly than released stops, their accuracy in detecting unreleased stops was greatly reduced. Released stops (phonotactically illegal in Korean) were more accurately detected than unreleased stops (phonotactically legal in Korean). This suggests that the acoustic-phonetic cues in the release component of the stops boosted detection accuracy, in spite of the illegality of coda releases in Korean. This is in line with the view that the more cues to segment identity in the speech signal, the more efficiently the segment is processed.

On the basis of the results of these two experiments, we propose that perception of non-native speech is influenced both by the phonotactics of the native language and by the richness of acoustic cues in the speech signal. Phoneme recognition is improved in a speech environment which is phonotactically legitimate in the native language, and, at the same time, any acoustic-phonetic cues present in the signal can be exploited by listeners, even if those cues are not the ones used in processing the listener's native language.

The interaction between phonotactics and acoustic-phonetic cues in non-native perception, however, appears to be further constrained by the language factor (Dutch vs. English stimuli). In Exp. 2a with Dutch stimuli, Korean listeners showed a phonotactic effect, but no effect of phonetic cue strength. It was only the English stimuli in Exp. 2b that brought about both effects. Given the assumption that the speech materials used here reflect a language difference rather than a speaker difference, the asymmetry between the Dutch and English stimuli may be due to differential familiarity with the non-native speech that is being processed.

Dutch is truly an 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. The language-dependent asymmetry in the results may then be ascribable to the degree of exposure to, or familiarity with, the spoken materials that are being processed. In the course of second language acquisition, Korean 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 phonotactics of Korean in English released stops may have led to slow detection times, familiarity with the form of such stops may have led to more accurate detection than for the unreleased English stops. In contrast, when Korean listeners heard stops which were both illegal and unfamiliar (Dutch released stops), target detection was no better than for the unreleased stops. There was thus only a phonotactic effect (in the RTs).

4. Summary and Conclusion

In the present study we used the phoneme monitoring task to explore how listeners of two unrelated languages, Dutch and Korean, process sounds which are phonotactically legal or illegal in either Dutch or Korean, when spoken by Dutch and English speakers. In Exp. 1, Dutch listeners' performance on

syllable-final stops in both Dutch and English was better when the stops were released than when they were unreleased. The results were consistent both with the hypothesis that native phonotactics plays a role in both native and non-native listening and with the hypothesis that released stops are informationally richer—the released coda stops were in harmony with the phonotactics of Dutch and, at the same time, carried more acoustic-phonetic cues than the unreleased coda stops. In Exp. 2, we found that native listeners of Korean, in which only unreleased coda stops are phonotactically viable, processed the unreleased stops more rapidly than the released stops. However, the error data with English but not with Dutch stimuli revealed that acoustic-phonetic cues in the signal were also exploited by Korean listeners: Released English stops were detected more accurately than unreleased ones. We proposed that, in non-native speech perception, phonotactic legitimacy in the native language speeds up phoneme recognition, that stronger acoustic-phonetic cues can improve recognition accuracy, and that familiarity with the non-native language modulates the relative influence of these two factors.

This study suggests that the mechanisms underlying speech perception cannot be fully understood without taking into account both the abstract phonological system of the listener's native language and the richness of the acoustic-phonetic cues in the speech signal. Furthermore, this study illustrates that cross-linguistic investigations of native and non-native listening can provide a valuable window on how experience with both native and non-native spoken language serves to shape the operation of the speech perception system.

5. Acknowledgements

We thank Kee-Ho Kim at Korea University for helping us with the acquisition of the Korean data. We also thank Anne Cutler, Elizabeth Johnson and Roel Smits for their comments and encouragement at various stages of this project.

6. References

- [1] Jusczyk, P.W. *The discovery of spoken language*. Cambridge, MA: MIT Press, 1997.
- [2] Werker, J.F. & Lalonde, C.E. Cross-language speech perception: Initial capabilities and developmental change. *Developmental Psychology*, 24: 672-683, 1988.
- [3] Cutler, A. Listening to a second language through the ears of a first. *Interpreting*, 5;1, 1-23, 2000/2001.
- [4] Best, C.T., & Strange, W. Effects of phonological and phonetic factors on cross-language perception of approximants. *J. of Phonetics*, 20: 305-330, 1992.
- [5] Costa, A., Cutler, A., Sebastián-Gallés, N. Effects of phoneme repertoire on phoneme decision. *Perception & Psychophysics*, 60:1022-1021, 1998.
- [6] Kuhl, P.K. & Iverson, P. Linguistic experience and the perceptual magnet effect, in W. Strange (Ed.) *Speech perception and linguistic experience*. Timonium, MD: York Press, 1995.
- [7] Weber, A. C. *Language-specific listening: The case of phonetic sequences*. Ph.D. dissertation, University of Nijmegen, MPI Series in Psycholinguistics, 16, 2001.
- [8] Hallé, P.A., Best, C.T. & Levitt, A. Phonetic vs. phonological influences on French listeners' perception of American English approximants. *J. of Phonetics*, 27: 281-306, 1999.