Journal of Phonetics 40 (2012) 443–452

Contents lists available at SciVerse ScienceDirect

Journal of Phonetics

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

Phonetic richness can outweigh prosodically-driven phonological knowledge when learning words in an artificial language

Sahyang Kim [a], Taehong Cho [b,*], James M. McQueen [c,d,e]

[a] Department of English Education, Hongik University, Seoul, Republic of Korea
[b] Hanyang Phonetics and Psycholinguistics Laboratory, Department of English Language and Literature, Hanyang University, Seoul 133-791, Republic of Korea
[c] Behavioural Science Institute, Radboud University Nijmegen, Nijmegen, The Netherlands
[d] Donders Institute for Brain, Cognition and Behaviour, Centre for Cognition, Radboud University Nijmegen, Nijmegen, The Netherlands
[e] Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands

ARTICLE INFO

Article history:
Received 2 February 2011
Received in revised form
19 January 2012
Accepted 14 February 2012
Available online 2 March 2012

ABSTRACT

How do Dutch and Korean listeners use acoustic–phonetic information when learning words in an artificial language? Dutch has a voiceless 'unaspirated' stop, produced with shortened Voice Onset Time (VOT) in prosodic strengthening environments (e.g., in domain-initial position and under prominence), enhancing the feature {−spread glottis}; Korean has a voiceless 'aspirated' stop produced with lengthened VOT in similar environments, enhancing the feature {+spread glottis}. Given this cross-linguistic difference, two competing hypotheses were tested. The phonological-superiority hypothesis predicts that Dutch and Korean listeners should utilize shortened and lengthened VOTs, respectively, as cues in artificial-language segmentation. The phonetic-superiority hypothesis predicts that both groups should take advantage of the phonetic richness of longer VOTs (i.e., their enhanced auditory–perceptual robustness). Dutch and Korean listeners learned the words of an artificial language better when word-initial stops had longer VOTs than when they had shorter VOTs. It appears that language-specific phonological knowledge can be overridden by phonetic richness in processing an unfamiliar language. Listeners nonetheless performed better when the stimuli were based on the speech of their native languages, suggesting that the use of richer phonetic information was modulated by listeners' familiarity with the stimuli.

1. Introduction

Cross-linguistic studies have shown that the prosodic structure of spoken utterances is manifested in the speech signal not only by various suprasegmental features (e.g., pitch movement) but also by fine-grained phonetic strengthening of individual segments at prosodically important landmarks (see Cho, 2011, for review). For example, segments are lengthened at the end of prosodic constituents (e.g., Cooper & Paccia-Cooper, 1980; Klatt, 1975; Wightman, Shattuck-Hufnagel, Ostendorf, & Price, 1992). They are also strengthened under prominence in stressed and accented syllables (e.g., Cambier-Langeveld & Turk, 1999; Cho & Keating, 2009; de Jong, 1995, 2004; de Jong & Zawaydeh, 2002; Klatt, 1975; Lehiste, 1970). Strengthening also occurs at the beginning of prosodic domains (e.g., Cho & Keating, 2001; Cho & McQueen, 2005; Fougeron & Keating, 1997; Keating, Cho, Fougeron, & Hsu, 2003), including spatio-temporal expansion of initial segments, a phenomenon known as domain-initial strengthening. The present study focuses on this phenomenon, and asks, cross-linguistically, if and how prosodically-driven durational cues in domain-initial stop consonants are used in segmenting the words of an artificial language.

Segmental variations reflecting prosodic structure have been shown to influence speech processing. Listeners can use the acoustic correlates of prosodic structure to decode those structures, facilitating segmentation of words at prosodic boundaries (e.g., Christophe, Peperkamp, Pallier, Block, & Mehler, 2004). For instance, word- and phrase-final lengthening appears to facilitate segmentation of words not only in continuous native-language speech (Kim & Cho, 2009; Salverda, Dahan, & McQueen, 2003), but also in artificial-language speech streams (Bagou, Fougeron, & Frauenfelder, 2002; Kim, Broersma, & Cho, 2012; Saffran, Newport, & Aslin, 1996; Tyler & Cutler, 2009). Likewise, it is well-known that stress patterns (cued e.g., by longer duration of stressed syllables in English and Dutch) are exploited in lexical segmentation (Cutler & Norris, 1988; Quené, 1993; Sluijter & van Heuven, 1996). As for the role of domain-initial strengthening in

* Corresponding author. Tel.: +82 2 2220 0746; fax: +82 2 2220 0741.
E-mail addresses: sahyang@hongik.ac.kr (S. Kim), tcho@hanyang.ac.kr, Taehong.Cho@gmail.com (T. Cho), James.McQueen@mpi.nl (J.M. McQueen).

0095-4470/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
doi:10.1016/j.wocn.2012.02.005

speech recognition, Cho, McQueen, and Cox (2007) have shown that acoustic correlates of domain-initial strengthening in English (such as longer VOT for voiceless stops and longer frication duration) can facilitate lexical segmentation. This suggests that detection of a possible lexical boundary is reinforced when it is aligned with a prosodic phrase boundary signaled by domain-initial strengthening cues. Building on this view, we ask here whether domain-initial cues are used in the segmentation of an artificial language. The cross-linguistic comparison of two typologically different languages, Dutch and Korean, allowed us to ask whether the use of the domain-initial strengthening cues in segmentation is modulated by the phonology of the listeners' native language.

We focused on Dutch and Korean voiceless stops. Dutch employs a binary phonological distinction in voicing between [−voice] and [+voice]. Unlike many other Germanic languages, a voiceless stop with [−voice] in Dutch is phonetically realized as unaspirated with relatively short Voice Onset Times (VOTs), while a voiced stop with [+voice] is produced with prevoicing (voicing lead), generating glottal pulsing during the stop closure (Slis & Cohen, 1969; Van Alphen & Smits, 2004). Although VOT appears to be the primary cue to the Dutch voicing contrast, other cues also signal the distinction (e.g., F0 and spectral center of gravity; Van Alphen & Smits, 2004). Korean, on the other hand, has a three-way stop contrast, and all three types are voiceless. These stops are often described as fortis (e.g., /p*/ with short VOTs), lenis (e.g., /p/ with intermediate VOTs) and aspirated (e.g., /p^h/ with long VOTs). Note again that other cues besides VOT signal the Korean three-way distinction (e.g., higher F0 for fortis and aspirated stops; creakiness on the following vowel for the fortis, and breathiness on the vowel for the lenis; Cho, Jun, & Ladefoged, 2002). Dutch and Korean voiceless stops thus differ in various ways on both phonetic and phonological dimensions. Critically, they also differ with respect to how domain-initial strengthening is applied to them.

In Dutch, the VOTs of the (unaspirated) voiceless stops are shorter at the beginning of higher prosodic domains (e.g., Intonational Phrases) than at the beginning of lower prosodic domains (e.g., Prosodic Words), and are shorter under prominence (Cho & McQueen, 2005). In contrast, in Korean (as in English), voiceless aspirated stops have longer VOTs at the beginning of higher prosodic domains and under prominence (e.g., Cho & Keating, 2001, 2009; Cho, Lee, & Kim, 2011; Keating et al., 2003; Pierrehumbert & Talkin, 1992).

Cho and McQueen (2005) argued, in line with Keating (1984), that while voiceless stops can be phonologically specified with [−voice] across languages, their phonetic realization is language-specific. Languages with voiceless unaspirated stops such as Dutch enhance the phonetic feature {−spread glottis} in prosodic strengthening environments, resulting in shorter VOTs. (Note that the square brackets '[ ]' refer to a phonological feature, and the curly brackets '{ }' to a phonetic feature, following Keating, 1984). In languages with voiceless aspirated stops (e.g., Korean), however, it is the phonetic feature {+spread glottis} for aspiration that is phonetically enhanced in prosodic strengthening environments, resulting in longer VOTs. Language-specific phonetic realizations of prosodic strengthening, especially under prominence, have also been examined by de Jong (1995, 2004), de Jong and Zawaydeh (2002), and Silbert and de Jong (2008). They found that the durational difference between vowels before voiced versus voiceless consonants (i.e., longer vowels before voiced consonants) is enhanced under stress in English but not in Arabic. Based on the findings, they claim that such voicing-induced durational differences are "linguistically specified" in English but not in Arabic, and that only linguistically-specified phonetic content is strengthened under prominence. de Jong and colleagues thus suggest that hyper-articulation is modulated by the phonological structure of a given language. They further propose that prominence (especially focus) can be used as a diagnostic for the phonetic content that is linguistically specified in a given language, and that communicative effectiveness is achieved by hyper-articulating the specified phonetic content. This account allowed us to predict that the language-specific phonetic enhancement that arises in prosodically strong environments is likely to be made use of by listeners in speech comprehension, regardless of the enhancement's specific phonetic manifestation. We tested this by asking Dutch and Korean listeners to listen to an unsegmented speech stream in an artificial language, and by then testing them on how well they were able to segment and learn the words of that language. If language-specific domain-initial strengthening cues are used in this segmentation task, Dutch listeners should learn words better if word-initial stops have shorter VOTs, while Korean listeners should benefit from longer VOTs.

For Dutch listeners, however, the use of language-specific phonological knowledge favoring stops with shorter VOTs would have to take place in spite of the fact that stops with longer VOTs carry richer acoustic–phonetic information with a greater auditory impact. That is, temporal expansion of the release burst and aspiration noise should enhance the percept of a voiceless stop at an auditory level not only through greater temporal separation between the closure and the voicing onset (Summerfield & Haggard, 1974) but also through increasing the rate of auditory nerve firing after a silent period of stop closure (Delgutte, 1982; Delgutte & Kiang, 1984). Lengthened VOT with a greater degree of aspiration thus serves as an important auditory cue to the voicelessness of a stop (e.g., Repp 1979; see Wright, 2004, for a survey of auditory phonetic cues). We refer to these perceptual enhancements as arising from the greater *phonetic richness* of the longer stops. Note that although the manipulation here concerns temporal expansion, the notion of phonetic richness can be extended to cases where multiple acoustic–phonetic cues may work together to boost the auditory impact of a segment (e.g., the presence of both release and formant transitional cues for stop identity versus the presence of only one of those cues; Cho & McQueen, 2006). Some evidence about the auditory–perceptual effect of long VOTs can be found with perceptual data from listeners of languages such as Spanish and Polish, which have voiced versus voiceless unaspirated stops, as in Dutch. Spanish and Polish listeners not only are sensitive to changes of VOT in the short-lag region, which is crucial to the phonemic distinctions of their native languages, but are also sensitive to VOT differences in the long-lag region (Abramson & Lisker, 1973; Keating, Mikos, & Ganong, 1981). These studies suggest that stops with increased VOTs might have cross-linguistic auditory perceptual robustness.

It is not obvious that phonetic richness will dominate, however. It has previously been shown, for instance, that listeners do not necessarily use acoustic-phonetically richer information in processing an unfamiliar language—at least if the information is not employed by the phonological system of the listener's native language. Cho and McQueen's (2006) phoneme monitoring study, for example, found that Korean listeners, who are always exposed to unreleased word-final stops in their native language, detected unreleased stops (phonologically viable in their native language yet phonetically poorer) more efficiently than released stops (phonologically unviable in Korean yet phonetically richer) in processing speech in Dutch and English, while Dutch listeners showed the opposite pattern. Recent studies which employed an artificial language learning paradigm have also shown that listeners are not able to use an acoustically salient cue (i.e., high pitch) in lexical segmentation when the cue appeared in a

position where listeners would not normally expect that cue (Kim et al., 2012; Tyler & Cutler, 2009).

In summary, therefore, the question is whether Dutch listeners will find segmentation of an artificial language easier when VOTs of word-initial stops are shorter (in line with the effects of domain-initial strengthening in Dutch) or longer (because longer stops are phonetically richer). We thus contrasted the predictions of a *phonological-superiority hypothesis* and a *phonetic-superiority hypothesis*. The phonological-superiority hypothesis predicts that, in segmenting an unfamiliar artificial language, listeners will make use of the acoustic–phonetic patterns of segments that are matched phonologically with those in prosodically strong environments in their native language, even if the phonetic information is less robust than that which occurs in prosodically weak environments. Under this hypothesis, Dutch listeners are expected to make more use of shortened VOTs than of lengthened VOTs as facilitative cues in artificial-language segmentation. It should be noted, however, that the Dutch data reported by Cho and McQueen (2005) did not provide direct comparisons of word-initial versus word-medial stops in phrase-internal contexts. We therefore do not know whether word-initial stops embedded in a phrase are indeed produced with shorter VOTs than word-medial stops in Dutch. But we do know that phrase-boundary cues (which necessarily also occur at word boundaries) facilitate lexical segmentation relative to phrase-internal word-boundary cues (e.g., Cho et al., 2007; Christophe et al., 2004; Kim & Cho, 2009). Given that shortened VOTs mark Dutch phrase boundaries, Dutch listeners may make use of these cues when segmenting novel lexical sequences that occur in an artificial language.

Alternatively, the phonetic-superiority hypothesis predicts that Dutch listeners will benefit from stops with lengthened VOTs, because longer stops are phonetically richer, with temporally expanded burst noise and aspiration that enhances the percept of the voicelessness of the stop at the auditory level. Longer VOTs should thus heighten the phonetic clarity of the word onsets, making it easier to segment the artificial language. Under this hypothesis, Dutch listeners are therefore expected to learn new words in an unfamiliar artificial language better when the words start with stops with longer VOTs.

Note that, for the Korean listeners, the two hypotheses predict the same result. Koreans should find long VOT more useful than short VOT in lexical segmentation either because lengthened VOTs are observed in prosodically strong environments (Cho et al., 2011), in line with the phonological-superiority hypothesis, or because lengthened VOTs are acoustic-phonetically more robust, in line with the phonetic-superiority hypothesis. The Korean listeners therefore constitute a control group.

2. Experiment

We tested the phonological- and phonetic-superiority hypotheses using an artificial language learning paradigm. Participants (native speakers of either Dutch or Korean) first listened to a 20-min pauseless stream of novel trisyllabic words from an artificial language (henceforth AL). After this learning phase, they were tested on whether they had learned the words that constituted the AL in a forced-choice identification task. In this test phase, pairs of trisyllabic sequences were presented (one was an actual word used in the AL, i.e., a pattern that had recurred in the training sequence, and the other was a nonword that was not a recurring pattern). Listeners had to identify the words in the test phase. This paradigm has been shown to be effective in testing effects of specific cues in the speech signal without recourse to any prior lexical knowledge and it has also revealed listeners' use of native segmentation cues in lexical segmentation of these unfamiliar, non-native languages (e.g., Bagou et al., 2002; Kim et al., 2012; Saffran et al., 1996; Tyler & Cutler, 2009).

We tested Dutch listeners to evaluate the hypotheses, and Korean listeners as a control group. We used the same speech materials with both listener groups. Further, we prepared two matched sets of speech materials, one based on the voice of a Dutch speaker (the Dutch-voice condition) and the other based on the voice of a Korean speaker (the Korean-voice condition). This allowed us to observe potential familiarity effects arising from differences in the pronunciation of the materials by Dutch and Korean speakers—that is, whether listeners would benefit when the spoken stimuli were created based on their native language. Although the same AL was used in all conditions (i.e., it always had the same lexical items) and long and short VOTs were presented in both Dutch- and Korean-voice conditions, the actual VOT manipulation (i.e., the short versus long VOT values that were used) therefore did differ between the language conditions, in order to reflect the difference in the natural VOT range across languages. Short and long VOT values appropriate for Dutch and Korean materials were selected in a phonetic-categorization pretest (see Table 2 for actual VOT values used in the study).

Finally, we tested the listeners' ability to identify not only the stop-initial words in the AL, but also nasal-initial words. Only the stop-initial words underwent the VOT manipulation, but successful segmentation of the words in the learning phase should result not only in the ability to recognize the manipulated (stop-initial) words but also the other (nasal-initial) words in the continuous stream. Testing the nasal-initial words thus allowed us to establish whether there was an across-the-board learning effect or whether learning was restricted to the stop-initial words.

In summary, we asked a number of questions. First, do listeners use differences in VOT to assist in segmenting the words of an artificial language? If so, there should be a difference in the number of words correctly identified between the short- and long-VOT conditions. Second, do listeners use the phonology of their native language in determining how they use VOT in segmenting an artificial language? If so, Dutch listeners should identify more words correctly in the short- than in the long-VOT condition, while Korean listeners should do the reverse. Third, alternatively, do listeners segment an artificial language more successfully when there are phonetically-richer cues to word-initial stops? If so, both Dutch and Korean listeners should identify more words correctly in the long- than in the short-VOT condition.

2.1. Method

2.1.1. Participants

In the Dutch-voice conditions, there were 16 Dutch and 30 Korean participants in each VOT condition (i.e., the short- and long-VOT conditions), and hence a total of 92 participants. In the Korean-voice conditions, there were 16 Korean and 30 Dutch participants in each VOT condition, again resulting in 92 in total. This between-subject design is illustrated in Table 1. Note that we started with 16 listeners in each condition, but initial analyses of the data showed that listeners who were exposed to non-native speech materials often failed to learn the AL, contributing to greater variability than in the native-speech conditions. We therefore increased the number of participants in the conditions where listeners were exposed to non-native speech materials to 30 per condition. Participants were assigned at random to either the short- or the long-VOT condition. Dutch participants were recruited from Radboud University Nijmegen, The Netherlands, and Korean participants from Hanyang University in Seoul, Korea. All of them were university students with normal hearing.

Table 1
Experimental design.

Listener language	Stimulus language	VOT condition	Number of participants
Dutch	Dutch (native speech)	Long	16
		Short	16
	Korean (non-native speech)	Long	30
		Short	30
Korean	Korean (native speech)	Long	16
		Short	16
	Dutch (non-native speech)	Long	30
		Short	30

They were all paid for their participation. Since the experiment was expected to be difficult and boring, we informed the participants before each experimental session started that there would be monetary reward for participants who scored 70% or more correct in the test.

2.1.2. Materials

An AL with six trisyllabic words was created. Five consonants (/p/,/t/,/k/,/m/,/n/) and three vowels (/a/,/i/,/u/) which exist in the Korean and the Dutch phoneme inventories were selected. These eight segments were combined to make 15 distinct CV syllables, which were further combined to make six trisyllabic words: [kanipu], [tumita], [pimaki], [namiku], [nutipa], [mutani]. Three of the six words had word-initial oral stops and three had word-initial nasals. None of them were existing words in Korean or Dutch.

The stimuli were created by concatenating syllables produced by a female native speaker of Dutch (for the Dutch-voice condition) and a male native speaker of Korean (for the Korean-voice condition). Both speakers were naïve as to the purpose of the experiment. The speakers produced all fifteen syllables 10 times both in isolation and in a CVCVCV context which was produced like a reiterant trisyllabic word with repeating syllables (e.g., [kakaka], [pipipi], etc.).

The same procedure was followed for the construction of the Dutch and Korean materials. In each case, we first selected, from the syllables produced in isolation, the most clearly articulated versions of those syllables. The acoustic values of these selected syllables were then manipulated such that the durations of the VOTs for all non-initial oral stops, the nasal murmurs for all nasals, and the vowels were normalized to the average values taken from the corresponding syllables in the medial position of the reiterant trisyllabic words (i.e., CV**CV**CV). For example, for a given test word [kanipu], the duration of each vowel (/a/,/i/,/u/) was set to the duration of that vowel in word-medial position in the appropriate reiterant trisyllabic word (e.g., [a] in [ka**ka**ka]), based on the average values of 10 repetitions of that vowel. The duration of the nasal murmur of [n] in the second syllable and the VOT of /p/ in the third syllable of [kanipu] were also based on the average values of [n] and [p] occurring in medial position in the appropriate reiterant trisyllabic words ([ni**ni**ni] and [pu**pu**pu], respectively). In this way, durations of all non-initial segments in the words of the AL were appropriate for those in word-medial position, in order to ensure that no robust word-initial or word-final cues existed in the materials except for the VOT manipulation of the initial stops.

Given that F0 rise in the following vowel may enhance the percept of voicelessness of the stop (Kingston & Diehl, 1994), and the fact that F0 serves as a cue to stops in Korean and Dutch (Cho et al., 2002; Van Alphen & Smits, 2004), it was also important to neutralize F0 across the speech streams. The average F0 values of all materials produced by each speaker were always used. After these normalization procedures, no stressed syllables could be identified. Adjustment of duration and F0 was done using the PSOLA function in Praat (Boersma & Weenink, 2011). These steps, admittedly, made the spoken artificial language sound somewhat unnatural, but they were necessary to ensure that all the syllables in the AL were free from any potential coarticulation or word-final lengthening effects, and from all word-initial strengthening effects except for the critical VOT manipulation on the word-initial stop consonants.

The VOT values of the word-initial stop consonants were manipulated to create the short- and long-VOT conditions for the Dutch and Korean materials. The VOT values for the short-VOT conditions were determined in language-specific ways based on the results of the pretest (see Section 2.1.3). For the long-VOT conditions, we simply added 55 ms to the short-VOT values to match the difference between VOT conditions across voice conditions. The VOT values are summarized in Table 2.

Note that the VOTs of the other stops used in the non-initial syllables in the Dutch- and Korean-voice conditions were 37 ms and 55 ms, respectively. These values are the average durations of the medial stops in the reiterant trisyllabic words produced by the Dutch and Korean speakers.

Finally, it was necessary to keep closure duration the same across the Dutch- and Korean-voice conditions and across the long and short VOT conditions, so that closure duration differences could not differentially signal lexical boundaries. We chose 55 ms as the stop closure duration, based on the average closure value of all the word-medial stops of the trisyllabic words (e.g., [kakaka], [pipipi]) produced by the two speakers. This is a rather arbitrary value because it is based on speakers of different languages, but it was reasonably short so that it was at least more appropriate as a word-internal stop closure than as a word-initial one. (Note that it would be interesting to see a possible cue-trading relationship between VOT and closure duration (cf. Repp, 1979), but here we focused on effects of VOT alone.)

The VOT manipulated word-initial syllables as well as the rest of the normalized syllables were concatenated to make six trisyllabic words. The words were then concatenated in random order without any pause between them to yield an approximately 10-min speech stream. One word never occurred twice in a row, and each word occurred 144 times. Transitional probabilities between syllables (i.e., the probability of the syllable sequence XY given the probability of the first syllable X) within words ranged from 0.5 to 1, and those across words ranged from 0.08 to 0.28.

2.1.3. Pretest: phonetic categorization

In order to make sure that we selected shortest VOT values that were still categorically perceived as the intended phonemes

Table 2
VOT values for word-initial syllables in the short- and long-VOT conditions.

Stop place	Short VOT (ms)	Long VOT (ms)
Dutch-voice conditions		
Bilabial	15	70
Alveolar	20	75
Velar	20	75
Korean-voice conditions		
Bilabial	50	105
Alveolar	35	90
Velar	55	110

(i.e., voiceless unaspirated for Dutch and aspirated for Korean) by native listeners, we conducted categorization tests on the syllables [ka], [pi], and [tu]—the syllables which were to be used as the onsets of the three stop-initial words in the AL.

2.1.3.1. Participants. Thirty-two Dutch listeners were recruited from Radboud University Nijmegen, The Netherlands, and thirty-two Korean listeners from Hanyang University in Seoul, Korea. None of the participants had hearing problems, none took part in the main experiment, and all were paid.

2.1.3.2. Materials. The syllables originally recorded and manipulated for the AL were concatenated to create disyllabic pretest materials. For both languages, consonants were tested in two separate disyllabic contexts: the target syllables (i.e., [ka], [pi], [tu]) were always the second syllables, and the initial syllables were either the identical syllables (e.g., [kaka], CV**CV**) or a syllable composed of a bilabial nasal stop and the same vowel as the second syllable (e.g., [maka], NV**CV**). The target syllables were put into these contexts in order to match them with the speech materials to be used in the main experiment. (Note that, of the 32 participants in each language group, 16 were presented with the CVCV context and the other 16 with the NVCV context.)

For the Dutch experiment, there were 19 VOT steps: four negative (prevoiced) VOTs (−100 ms, −75 ms, −50 ms, −25 ms) and 15 positive (voiceless) VOTs (from 0 to 70 ms, 5 ms apart). (Note that only four steps were used with prevoicing because Dutch listeners are not sensitive to variation in negative VOTs; Van Alphen & McQueen, 2006.) In order to create these tokens, VOT portions were first taken from one representative voiced token (for negative VOTs) and one voiceless token (for positive VOTs). Their durations were then manipulated using the PSOLA function of Praat to create different steps. The resulting VOT portions were concatenated to a matched vowel which was taken from another representative syllable containing that vowel. For the Korean experiment, different VOT values were used because Korean stops are all voiceless stops—i.e., fortis, lenis and aspirated stops (from the shortest to the longest VOT values). (See Cho et al., 2002 for other phonetic correlates of the three-way stop contrast in Korean.) So VOT values in Korean ranged from 5 ms to 90 ms and each step was 5 ms apart, and hence there were 18 steps. VOTs were manipulated in a similar way to the Dutch stimuli. The VOT values used were by and large within the range of the values observed across various conditions reported in previous production studies (see Cho & McQueen, 2005; Van Alphen & Smits, 2004, for Dutch, and Cho & Keating, 2001; Cho et al., 2002, for Korean).

2.1.3.3. Procedure. Dutch subjects participated in a 2AFC task (voiced or voiceless stops) and Korean subjects participated in a 3AFC task (fortis, lenis, or aspirated stops). The task was divided into two blocks. Tokens were pseudo-randomized in each block, and each token appeared 3 times in each block. Participants heard the auditory stimuli over headphones, and marked their answers (which stop they thought they had heard) on a response sheet.

2.1.3.4. Results. The results are summarized in Figs. 1 and 2. For the Dutch-voice condition, the shortest VOT was determined as the first value after the categorical perception boundary between voiced and voiceless stops with more than 80% voiceless stop responses (marked by a horizontal line in Fig. 1). For the Korean-voice condition, the shortest VOT was determined as the first value that Korean listeners categorized as an aspirated stop, again in at least 80% of trials (marked by a horizontal line in Fig. 2). Note that, while the Dutch categorization function has a steep slope,

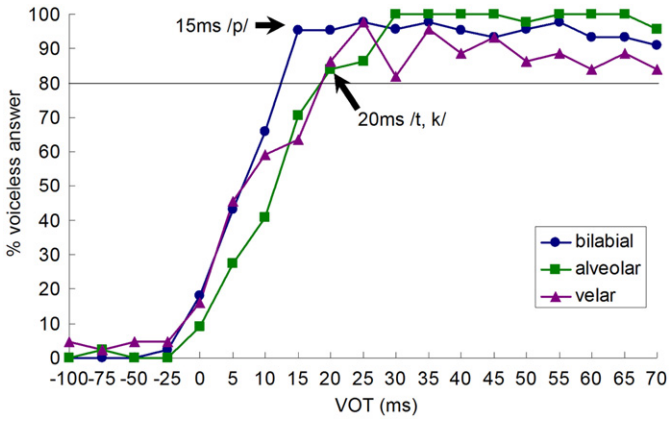

Fig. 1. Results of voiced–voiceless phonetic-categorization pretest in Dutch.

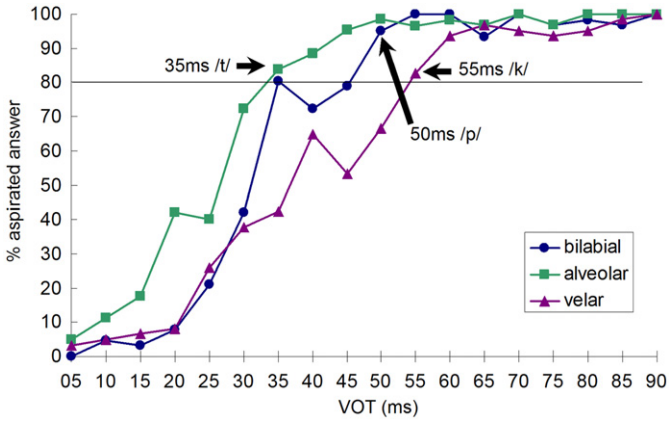

Fig. 2. Results of fortis–lenis–aspirated phonetic-categorization pretest in Korean.

that for Korean is less steep. There are at least two, possibly interrelated, reasons for this. First, Korean categorization was carried out in a 3AFC task (i.e., there were three categories that listeners had to choose from). Non-aspirated responses towards the left end of the VOT continuum in Fig. 2 are thus combined responses in the fortis and lenis categories (there was also no sharp boundary between fortis and lenis; not shown in the figure). Second, F0 was controlled to be flat in the vowels of the target-bearing sequences. Given that F0 is an important cue for aspirated–lenis and fortis–lenis distinctions in Korean, it is possible that listeners categorized aspirated versus lenis (and fortis versus lenis) less sharply.

Since place of articulation affects VOT values, we selected different VOT values for each initial consonant (see Figs. 1 and 2 and Table 2). Stops with these VOT values were used to make the Dutch- and Korean-voice AL materials, as described above.

Regarding the VOT values determined for word-initial versus word-medial position, it is worth pointing out that there is an apparent asymmetry between the short and long VOT conditions. In the Dutch-voice condition, for example, the short initial VOT (18.3 ms on average) differed from the non-initial VOT (37 ms) by 18.7 ms, while the difference between the long initial VOT (73.3 ms on average) and the non-initial VOT was 36.3 ms (there is a similar pattern in the Korean stimuli). The greater initial vs. medial contrast in VOT could potentially constitute a stronger segmentation cue in the long VOT condition than in the short VOT condition. There are several reasons, however, why this asymmetry does not pose a problem in interpreting the data. First, the asymmetry disappears if we consider the data in relative terms—i.e., the ratio of 18.3 ms to 37 ms in the Dutch short VOT

condition (0.49) is equivalent to the ratio of 37 ms to 73.3 ms in the Dutch long VOT condition (0.50). The ratio of VOTs may be more perceptually relevant than the absolute difference. Second, even if the absolute difference were also relevant, the stimuli in the Dutch short VOT condition were selected, via the pretest, to be good tokens of Dutch voiceless stops. They thus provide the optimal test of whether listeners apply knowledge of native language phonology in segmentation. Third, in creating the lexical items in the artificial language, it was important to maintain the VOT value of the medial stops across VOT conditions. If we had lengthened VOT in word-medial position (much longer than 37 ms) in the short VOT condition to match the absolute VOT difference in the long VOT condition, the variation in medial VOTs across conditions would make it impossible to interpret the data solely in terms of the role of initial VOTs. Sample speech files for the long and short VOT conditions in each voice condition are available in the electronic version of this paper as Supplementary Multimedia Data.

2.1.4. Procedure

Experimental sessions were composed of a learning phase and a test phase. During the learning phase, participants heard a speech stream from one of the VOT conditions (short or long). They were told that they would hear a speech stream from a simple AL which was composed of a series of nonsense words, and that there would be no pauses between words. They were informed that their task was to find the words of the AL from the speech stream. They were not told how many words were in the language. Each participant heard the concatenated sound stream played from a PC through headphones either at the Max Planck Institute for Psycholinguistics, Nijmegen, for the Dutch listeners, or at Hanyang University, Seoul, for the Korean listeners. They were asked to adjust the volume to the most comfortable level. Several participants were tested at the same time, whenever possible.

The learning session lasted approximately 20 min. Participants heard a 10-min speech stream, had a one-minute silent break, and then heard the same 10-min stream again. There were 20 ms fade-in and fade-out periods at the beginning and end of each speech stream, such that participants would not get any information about word boundaries from stream onsets and offsets.

In the test phase, there were 36 forced-choice pairs that were made from the combination of the six trisyllabic existing words of the AL and six trisyllabic novel strings. Three of these new strings were part-words and three were non-words, none of which existed in the AL. A part-word overlapped with an existing word, containing the final two syllables of the existing word plus an additional following syllable. The transitional probability between the overlapping string and the additional syllable was 0.22, which was less than the range of transitional probability within existing words (i.e., 0.5–1.0). Non-word strings were composed of the syllables that were used in the learning phase, but had sequences of syllables that had never been heard during the learning phase. Thus, the transitional probability of the non-words was zero.

On each trial, participants heard one forced-choice pair. There was an 800 ms inter-stimulus interval between the two members of a pair. All stimuli presented in the test phase had the same VOT values as those participants had been exposed to during the learning phase. Participants were given an answer sheet with the two alternatives in written form (in Roman alphabet for the Dutch, and in Hangul for the Koreans). They had to indicate which of the two words was part of the AL. They had four seconds to write an answer on each trial.

2.2. Results

2.2.1. Dutch-voice condition

A series of one-sample *t*-tests were first conducted to see whether listeners learned new words in the AL across initial consonant type (stop-initial versus nasal-initial words). Dutch listeners showed above-chance performance in the long VOT condition (67.6%, $t1(15)=5.25$, $p<0.001$; $t2(5)=3.61$, $p<0.05$), but not in the short VOT condition (50.5%, $t1(15)<1$, $t2(5)<1$). Korean listeners, on the other hand, performed at chance level in both VOT conditions (long VOT, 54.2%, $t1(29)=1.18$, $t2(5)=1.58$, both at $p>0.1$; short VOT, 47.1%, $t1(29)=1.18$, $t2(5)=1.08$, both at $p>0.1$).

We then examined interaction effects between VOT (short versus long) and Consonant Type (stop-initial versus nasal-initial words) in repeated measures ANOVAs in order to see whether the VOT factor would affect segmentation of only the stop-initial test words, which are expected to be directly influenced by VOT differences, or also of the nasal-initial words, which could potentially be segmented more easily after successful segmentation of neighboring stop-initial test words.

For Dutch listeners, in line with the results of the chance-level tests, there was a significant main effect of VOT both by subjects and items ($F1(1, 30)=16.740$, $p<0.001$, $F2(1, 4)=97.942$, $p=0.001$). More words were correctly identified in the long- than in the short-VOT condition (67.6% vs. 50.5%), as illustrated in Fig. 3a. There was also a significant main effect of Consonant Type only in the by-subject analysis ($F1(1, 30)=6.7$, $p=0.015$): there was a tendency towards Dutch listeners performing better on stop-initial than on nasal-initial words. However, there was an interaction effect between VOT and Consonant Type in the by-item analysis ($F2(1, 4)=13.21$, $p=0.022$). As can be seen from Fig. 3b and c, planned pairwise *t*-tests along with eta^2 statistics suggested that this trend interaction arose because, although the effect of VOT was significant with both stop-initial words and nasal-initial words, the effect was larger with oral stop-initial words (mean diff. 23.2%, $t1(15)=4.89$, $p<0.001$, $t2(2)=8.57$, $p<0.5$; eta$^2=0.974$) than with nasal-initial words (mean diff. 10.8%, $t1(15)=2.59$, $p<0.05$, $t2(2)=5.09$, $p<0.05$; eta$^2=0.929$).

For Korean listeners, there was a main effect of VOT, but only in the analysis by item ($F1(1, 58)=2.66$, $p>0.1$; $F2(1, 4)=26.66$, $p<0.01$), showing a tendency towards more accurate word identification in the long-VOT condition (54.2%) than in the short-VOT condition (47.1%), as shown in Fig. 4a. There was no Consonant Type effect ($F1(1, 58)=1.13$, $p>0.1$; $F2(1, 4)<1$,

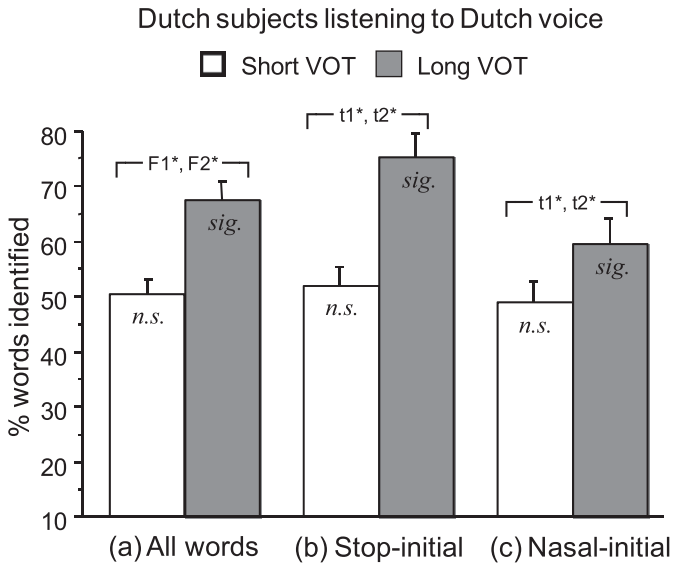

Fig. 3. % correct word identification for the short- vs. long-VOT conditions by Dutch listeners exposed to the Dutch-voice artificial language. * marks a statistically significant difference ($p<0.05$) between two conditions; *sig.* and *n.s.* indicate, respectively, whether performance in a given condition was significantly different from chance (50%), or not. (Error bars refer to standard errors.)

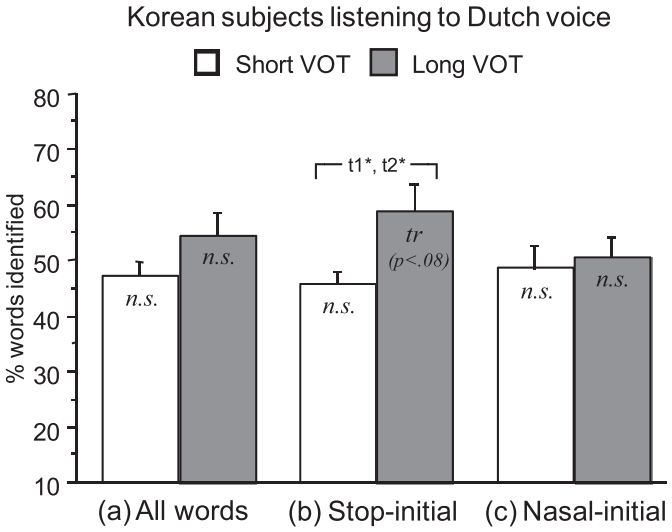

Fig. 4. % correct word identification for the short- vs. long-VOT conditions by Korean listeners exposed to the Dutch-voice artificial language. * marks a statistically significant difference ($p < 0.05$) between two conditions; *sig.* and *n.s.* indicate, respectively, whether performance in a given condition was significantly different from chance (50%), or not. (Error bars refer to standard errors.)

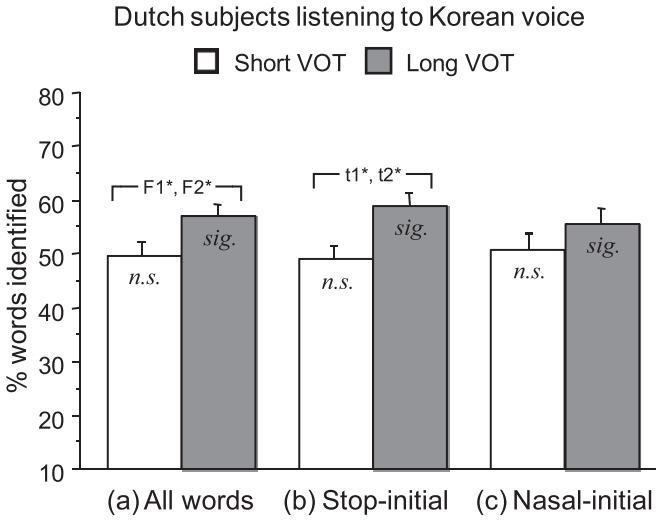

Fig. 5. % correct word identification for the short- vs. long-VOT conditions by Dutch listeners exposed to the Korean-voice artificial language. * marks a statistically significant difference ($p < 0.05$) between two conditions; *sig.* and *n.s.* indicate, respectively, whether performance in a given condition was significantly different from chance (50%), or not. (Error bars refer to standard errors.)

$p > 0.1$), but there was a robust VOT by Consonant Type interaction effect ($F1(1, 58) = 4.51$, $p < 0.05$; $F2(1, 4) = 15.01$, $p < 0.05$). As shown in Fig. 4b and c, planned pairwise comparisons revealed that a significant effect of VOT was found only with stop-initial words. More words were correctly recognized in the long- than in the short-VOT condition (58.7% vs. 45.7%; $t1(29) = 2.49$, $p < 0.05$; $t2(2) = 12.57$, $p < 0.01$). No such effect was observed with nasal-initial test words (48.5% vs. 50.4%, $t1(29) < 1$, $t2(2) < 1$).

In sum, when Dutch listeners were exposed to Dutch-voice materials, lengthened VOTs of the initial stop of the test words helped them to learn not only stop-initial test words themselves but also nasal-initial test words, showing an across-the-board learning effect, although the learning effect was more robust with stop-initial test words. When Korean listeners were exposed to Dutch-voice materials, the long VOT advantage was found only in the stop-initial test words.

2.2.2. *Korean-voice condition*

A series of one-sample *t*-tests revealed that both Dutch and Korean listeners performed above chance in the long VOT condition (Dutch listeners, 57.2%, $t1(29) = 3.44$, $p < 0.005$, $t2(5) = 2.92$, $p < 0.05$; Korean listeners, 66.8%, $t1(15) = 2.98$, $p < 0.01$, $t2(5) = 4.31$, $p < 0.01$). In the short VOT condition, both Dutch and Korean listener groups showed chance-level performance (Dutch listeners, 49.9%, $t1(29) < 1$, $t2(5) < 1$; Korean listeners, 52.5%, $t1(15) < 1$, $t2(5) < 1$).

For the Dutch listener group, repeated measures two-way ANOVAs (with VOT and Consonant Type factors) revealed a significant main effect of VOT ($F1(1, 58) = 5.36$, $F2(1, 4) = 15.04$, both at $p < 0.05$), showing overall better performance in the long-VOT condition (57.2%) than in the short-VOT condition (49.9%) (Fig. 5a). There was no effect of Consonant Type ($F1(1, 58) < 1$, $F2(1, 4) < 1$), nor was there a VOT by Consonant Type interaction ($F1(1, 58) = 1.14$, $p > 0.1$; $F2(1, 4) = 2.44$, $p > 0.1$). The null interaction effect suggests an across-the-board VOT effect—i.e., long VOTs in stop-initial test words helped listeners to learn not only stop-initial test words themselves, but also nasal-initial test words. However, planned pairwise comparisons revealed that the VOT effect was significant only in the stop-initial test word condition (see Fig. 5b; mean diff. 6.9%, $t1(29) = 2.52$, $t2(2) = 6.8$, both at $p < 0.05$). The nasal-initial test word condition showed no significant VOT effect, though the direction of the effect was maintained (see Fig. 5c; mean diff. 5%, $t1(29) = 1.21$, $t2(2) = 1.26$, both $p > 0.1$).

For the Korean listener group, there was again a significant main effect of VOT ($F1(1, 30) = 4.53$, $p < 0.05$; $F2(1, 4) = 32.02$, $p < 0.01$) (Fig. 6a). As was the case with the Dutch listener group, there was neither a Consonant Type effect ($F1(1, 30) < 1$, $F2(1, 4) < 1$) nor a VOT by Consonant Type interaction ($F1(1, 30) = 1.57$, $p > 0.1$; $F2(1, 4) = 3.39$, $p > 0.1$). However, as shown in Fig. 6b and c, planned pairwise comparisons showed that, unlike the Dutch listener group, the VOT advantage was observed in the nasal-initial condition, although the VOT effect was more robust in the stop-initial test word condition than in the nasal-initial test word condition—that is, the size of the VOT effect was larger in the stop-initial condition (mean diff. 19.9%, $t1(15) = 2.46$, $p < 0.05$, $t2(2) = 3.93$, $p < 0.06$) than in the nasal-initial condition (mean diff. 9.7%, $t1(15) = 1.28$, $p > 0.01$, $t2(2) = 6.42$, $p < 0.05$).

In sum, when Dutch listeners were exposed to the Korean-voice AL, they learned the AL quite efficiently when the stop-initial test words started with long VOTs (in the long VOT condition), while short VOTs did not help them to learn the AL. But the VOT effect did not appear across the board (there was a significant VOT effect only on the stop-initial test words), unlike in the Dutch-voice AL conditions (where there was a long-VOT advantage for both consonant types). In contrast, when Korean listeners processed the Korean-voice AL, they showed the long VOT advantage across consonant types. This pattern is also different from the results found with the Dutch-voice AL, where Korean listeners showed the long VOT advantage only in the stop-initial test word condition.

3. Discussion

The main goal of the present study was to examine how listeners of Dutch, which has voiceless unaspirated stops with relatively short VOTs, would use long versus short VOTs of word-initial stops in lexical segmentation of an unfamiliar (artificial) language, and if they would differ from listeners of Korean, which is typologically different, having aspirated stops produced with long VOTs.

Two competing hypotheses were considered. The phonological-superiority hypothesis predicted that Dutch listeners would

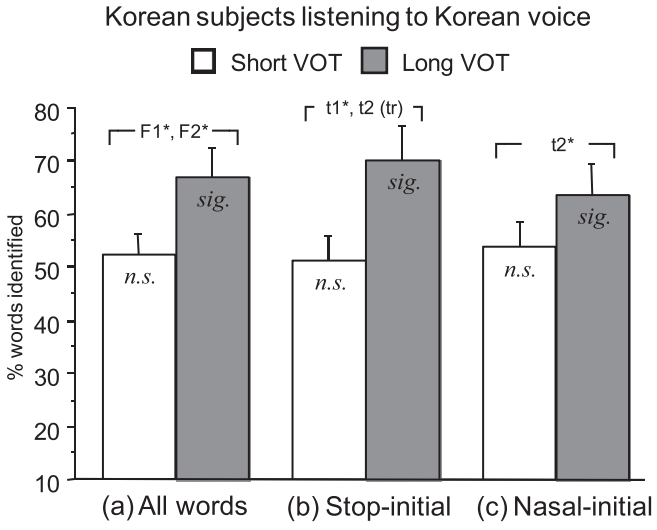

Fig. 6. % correct word identification for the short- vs. long-VOT conditions by Korean listeners exposed to the Korean-voice artificial language. * marks a statistically significant difference ($p < 0.05$) between two conditions; (tr) indicates a trend ($p < 0.06$); *sig.* and *n.s.* indicate, respectively, whether performance in a given condition was significantly different from chance (50%), or not. (Error bars refer to standard errors.)

make use of shortened VOTs of word-initial voiceless stops as a cue for lexical segmentation in line with the phonological aspects of the language: voiceless (unaspirated) stops are specified with its language-specific phonetic feature {−spread glottis} which is enhanced in domain-initial positions, resulting in shortened VOTs (Cho & McQueen, 2005). Given that the language-specific enhancement of phonetic content plays an important role in marking phonological and lexical contrast in the language (de Jong, 2004; de Jong & Zawaydeh, 2002), and given that phrase-initial strengthening cues of the language are used as a cue for lexical segmentation in the native language (Cho et al., 2007), shortened VOTs should signal phrase boundaries in the artificial language, making it easier to segment and learn the words of that language. Alternatively, under the phonetic-superiority hypothesis, it was predicted that Dutch listeners would make use of temporally expanded VOTs of word-initial voiceless stops in the segmentation of the artificial language. Given that lengthened VOTs carry richer acoustic–phonetic information specifying voicelessness, listeners could take advantage of this enhancement of the phonetic clarity of word-initial stops, and hence be able to segment the artificial language better.

Unlike Dutch listeners, however, the control Korean listeners were expected to make use of lengthened VOTs as a cue for lexical segmentation as this follows from both the phonetic- and the phonological-superiority hypotheses. VOTs for aspirated stops in Korean become lengthened in prosodically strong environments, which enhances both the phonetic richness of the sound and the {+spread glottis} feature, hence maximizing the phonological distinction between aspirated stops and other stops (Cho & Jun, 2000; Cho & Keating, 2001; Cho et al., 2011). Furthermore, Korean listeners have previously been found to be able to use the acoustic correlates of domain-initial strengthening (which include longer VOT) in processing English as an L2 (Kim & Cho, 2010). Not surprisingly, therefore, we found here that Korean listeners made use of lengthened VOTs to improve segmentation of the artificial language. Crucially, however, the Dutch listeners also made use of lengthened VOTs, and not shortened VOTs, to improve their segmentation performance. This was true not only in the Korean-voice condition, in which the long VOTs (101.6 ms) were excessively long as tokens of Dutch voiceless stops, but also in the Dutch-voice condition, where the long VOTs were less extreme (73.3 ms), and the speech materials were recorded by a Dutch speaker. This result suggests that Dutch listeners do take advantage of the phonetic richness associated with longer VOTs in processing an unfamiliar language, in spite of the phonologically important role of shortened VOTs in their native language. The Dutch listeners' behavior therefore provides support for the phonetic-superiority hypothesis, and contradicts the phonological-superiority hypothesis.

Another question that the present study aimed to answer was whether listeners would benefit when the speech materials of the unfamiliar artificial language were created based on their native language—that is, with the stimuli recorded by a speaker of their native language, and with the long versus short VOT values determined by the phonetic categorization of native listeners. The results revealed that both Korean and Dutch listener groups indeed performed better when they processed native-language-based materials. First, listeners had more difficulty learning the artificial languages based on non-native materials (hence the larger numbers of participants in these conditions). Second, mirror-image patterns were observed in the test results. In processing Dutch-based stimuli, Dutch listeners' learning performance was better in the long VOT condition than in the short VOT condition, not only with the stop-initial test words that were directly influenced by long VOTs, but also with nasal-initial test words (i.e., there was a robust across-the-board learning effect). In processing the same Dutch-based stimuli, Korean listeners also performed better in the long-VOT than in the short-VOT condition, but the learning effect was not as robust as that shown by the Dutch. The learning effect in the Korean group was limited to stop-initial test words, and the effect size on learning stop-initial words was far smaller compared to that in the Dutch group (58.7% vs. 75.4%). Crucially, however, the exact mirror-image pattern was observed with Korean-based stimuli. This time, Korean listeners showed the robust long VOT-induced learning effect on both stop-initial and nasal-initial test words, while the learning effect for the Dutch listeners was limited to stop-initial test words, whose effect size was again smaller than that of stop-initial test words by Korean listeners (58.7% vs. 70.1%). The mirror-image pattern across listener groups suggests that listeners benefit from familiar speech sounds in processing an unfamiliar language.

We suggest that these results indicate that, with respect to lexical segmentation of an artificial language, phonetic richness outweighs prosodically-driven phonological knowledge. The use of phonetic information, however, appears to be modulated by listeners' familiarity with the speech material. Listeners could extract more acoustic–phonetic information from speech signals when the speech materials are familiar to them (i.e., when they were listening to a speaker of their native language). This effect is robust enough, for Dutch listeners, to offset a potential advantage that might come from the longer VOTs used in the Korean-based stimuli. Understanding exactly which acoustic dimensions of the native speech material listeners benefited from is beyond the scope of the present study. Nevertheless, it seems likely that the precise realizations of the segments spoken by the two speakers (e.g., formant structures of the vowels, consonant-to-vowel formant transitions, and points of articulation of consonants) may have mapped more closely onto native-language vowel and consonant prototypes when those segments were spoken by a speaker of the participants' native language.

It should be noted, however, that the long-VOT advantage observed with Dutch listeners may not be entirely due to the phonetic richness of the speech signal. The VOT values employed in the long-VOT conditions deviated from the permissible range in Dutch, so the unnaturally long VOTs could make those stops perceptually stand out. This possibility is also in line with the assumption that a non-native sound category is learned better

when the perceived phonetic dissimilarity between the non-native sound and the closest native sound becomes greater (Best, 1995; Flege, 1995). It therefore remains to be seen how much the perceptual advantage of long VOT observed with Dutch listeners is attributable to its intrinsic auditory–perceptual robustness, and how much it is to its unnaturalness arising with deviation from its phonetic distribution of the native language.

However, lengthened VOTs indeed appear to carry cross-linguistically applicable auditory–perceptual robustness. Previous studies have suggested that native listeners of languages with voiceless unaspirated stops are also sensitive to differences in longer VOTs. In Abramson and Lisker (1973), for example, Spanish listeners perceived a phonemic boundary appropriate for the phonological contrast in their native language (voiced vs. voiceless unaspirated), but they were also able to discriminate stop variants with longer VOTs, which Abramson and Lisker attributed to the special psychoacoustic status of long VOTs. Similarly, Keating et al. (1981) showed that native listeners of Polish (which also has voiced vs. voiceless unaspirated stops) showed sensitivity to the change of VOT range including longer VOTs, even when the range is not used in Polish. Most recently, Broersma (2009) demonstrated that Dutch listeners who were not trained in Korean could nevertheless discriminate the Korean three-way stop contrast. That is, just like Koreans, Dutch listeners were able to distinguish between fortis and aspirated stops, with VOTs more or less equivalent to voiceless unaspirated stops in Dutch and voiceless aspirated stops in English, respectively.

The common feature of these previous studies and the present one is that listeners of the languages which do not use long VOT cues in their phonological system are still sensitive to the phonetic cue with longer VOTs. They all have only voiceless unaspirated stops, so that longer VOTs along the positive VOT continuum are not used by the phonology of these languages (Dutch, Spanish, and Polish). It is thus also plausible that the phonetic richness of speech sounds in an unfamiliar language is most effectively exploited when it does not interfere with the phonological system of the native language—that is, if the cue is phonetically implemented in a part of the acoustic–phonetic space which is not used by the phonology of the listener's native language. In the present case, lengthened VOTs strongly signaled voiceless stops to the Dutch listeners (and no other phonemes). This richer marking of the stops made it easier to segment words beginning with those stops out of the artificial speech stream. We therefore propose that while one cannot entirely rule out the possible involvement of the unnaturalness of long VOTs in Dutch listeners' learning an AL, the intrinsic auditory–perceptual robustness of long VOTs plays a role in learning an unfamiliar language.

Our findings also have other implications for the roles of language-specific phonological knowledge and phonetic richness in processing an unfamiliar language. It has generally been agreed by researchers that adult listeners tend to perceive speech sounds through the 'phonological filter' of their native language, so that their sensitivity to contrastive phonetic differences is tuned according to their experience with the phonological systems of their native language (e.g., Best, 1995; Cutler & Broersma, 2005; Flege, 1995, among many others). In particular, as discussed in the introduction, the fact that Korean listeners perceive (phonetically poorer) unreleased stops better than (phonetically richer) released stops in processing non-native speech has provided a concrete example of a situation where phonological experience overrides phonetic richness in non-native speech perception (Cho & McQueen, 2006). The language-specificity of speech perception has also been observed in lexical segmentation of unfamiliar artificial languages or non-native languages—e.g., the use of language-specific phonotactics (Weber & Cutler, 2006) and rhythmic/prosodic cues (Cutler & Otake, 1994; Kim et al., 2012; Tyler & Cutler, 2009) in processing non-native speech. The results of the present study therefore appear to be in contrast to the generally observed dependency that listeners have on their phonological knowledge in the perception of unfamiliar or non-native speech.

A question then follows. Why are some phonological aspects of the native language strongly rooted in processing an unfamiliar language, but why not in other cases, like the one in the current study? While it is not yet possible to provide a definite answer to this question, several possibilities can be thought of. One possibility may have to do with the fact that the phonological shortening effect on VOT in prosodic strengthening environments in Dutch appears not to be very large (i.e., a significant but small effect with about 5 ms shortening of VOT, as reported by Cho & McQueen, 2005). Small phonetic effects of phonological enhancement may not be effectively transferred to processing an unfamiliar language. If this is the case, the manipulation of VOT alone employed in the present study may not be a good gage of potential perceptual effects of the actual production characteristics of prosodic variation. Since prosodically-driven VOT shortening is accompanied by lengthening of the stop closure duration and the following vowel (Cho & McQueen, 2005), VOT shortening may become perceptually relevant only when it works together with these other lengthening cues. Alternatively, prosodic strengthening effects that are perceptually relevant may be associated with the following vowel because various other prosodic cues (pitch, duration and amplitude) are available in the vowel. Further studies are warranted to explore these possibilities.

In the absence of an explanation for the lack of a VOT shortening effect in the present Dutch-listener data, any interpretation must therefore be treated with caution. Nevertheless, we wish to propose the following possibility: The degree of the listener's dependency on their phonological experience in processing an unfamiliar language may be modulated by the auditory–perceptual robustness of the segments involved—what we have referred to as phonetic richness. As discussed in the introduction, lengthened VOTs may have a greater auditory–perceptual impact, which is likely to be used cross-linguistically, presumably providing more information about the voicelessness of the stop. That is, the more informative a segment is, the more likely it is to be used by the listener in processing an unfamiliar language. Phonetic richness of lengthened VOTs could then sometimes be powerful enough to supersede possible perceptual effects of shortened VOTs that may exist in the listener's native phonology.

4. Conclusion

The present study showed that both Dutch and Korean listeners took advantage of long VOTs in segmenting words in an unfamiliar (artificial) language, regardless of whether lengthened VOT is used by the prosodic system of the listeners' native language. Furthermore, Dutch listeners did not make use of shortened VOTs, which are a phonetic consequence of enhancement of the language-specific {−spread glottis} feature under prosodic strengthening. Dutch listeners' failure to exploit shortened VOTs may be interpreted in different ways, but it nonetheless suggests that the effect of the listeners' phonological knowledge can be overridden by phonetic richness. While previous studies have emphasized listeners' dependency on language-specific phonological knowledge in non-native speech perception, it is proposed that phonetic richness of speech sounds may also play a role in speech processing under some circumstances—i.e., when the auditory–perceptual robustness of the non-native phonetic cue is powerful enough, and/or when the cue uses a part of the acoustic–phonetic space that is not used by the native language, so that it does not interfere with the phonetic realization of the phonological system of the language. It was also proposed that the interplay of phonetics and phonology in processing

an unfamiliar language is further modulated by listeners' familiarity with the acoustic–phonetic detail in that language. Our findings therefore suggest that the relative importance of phonetic and phonological factors vary across different listening situations. In the situation where the listener has to segment the words of a new language, application of phonological knowledge about how prosodic structure is phonetically implemented in the native language appears to be modulated by the physical dimension of phonetic richness. All in all, in order to achieve a better understanding of how listeners process unfamiliar or non-native speech, we need a unified model of non-native speech perception that does not focus solely on phonological aspects of the native and non-native languages, but also integrates the role of phonetic richness and its interplay with phonological and language-experience factors.

Acknowledgment

We wish to thank Korean and Dutch participants. We would also like to thank the three anonymous reviewers and the Editor for their constructive comments from which this paper has greatly benefited.

Appendix A. Supplementary materials

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.wocn.2012.02.005.

References

Abramson, A. S., & Lisker, L. (1973). Voice-timing perception in Spanish word-initial stops. *Journal of Phonetics*, *1*, 1–8.

Bagou, O., Fougeron, C., & Frauenfelder, U. H. (2002). Contribution of prosody to the segmentation and storage of "words" in the acquisition of a new mini-language. *Paper presented at Speech Prosody*. Aix-en-Provence, France.

Best, C. T. (1995). A direct realist perspective on cross-language speech perception. In: W. Strange (Ed.), *Speech perception and linguistic experience: Theoretical and methodological issues in cross-language speech research* (pp. 167–200). Baltimore, MD: York.

Broersma, M. (2009). Dutch listeners' perception of Korean stop triplets. *Poster presented at the Acoustical Society of America second special workshop on speech: Cross-language speech perception and linguistic experience*. Portland, OR, USA (Abstract appeared in *Journal of the Acoustical Society of America, 125*, 2775.).

Boersma, P., & Weenink, D. (2011). *Praat: Doing phonetics by computer [Computer program]*. Version 5.2.3.

Cambier-Langeveld, T., & Turk, A. (1999). A cross-linguistic study of accentual lengthening: Dutch vs. English. *Journal of Phonetics*, *27*, 255–280.

Cho, T. (2011). Laboratory phonology. In: N. C. Kula, B. Botma, & K. Nasukawa (Eds.), *The continuum companion to phonology*. London/New York: Continuum.

Cho, T., & Jun, S. (2000). Domain-initial strengthening as featural enhancement: Aerodynamic evidence from Korean. *Chicago Linguistics Society*, *36*, 31–44.

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

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

Cho, T., & Keating, P. A. (2009). Effects of initial position versus prominence in English. *Journal of Phonetics*, *37*, 466–485.

Cho, T., Lee, Y., & Kim, S. (2011). Communicatively-driven versus prosodically-driven hyper-articulation in Korean. *Journal of Phonetics*, *39*, 344–361.

Cho, T., & McQueen, J. M. (2005). Prosodic influences on consonant production in Dutch: Effects of prosodic boundaries, phrasal accent and lexical stress. *Journal of Phonetics*, *33*, 121–157.

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

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.

Cooper, W. E., & Paccia-Cooper, J. (1980). *Syntax and speech*. Cambridge, MA: Harvard University Press.

Cutler, A., & Broersma, M. (2005). Phonetic precision in listening. In: W. J. Hardcastle, & J. M. Beck (Eds.), *A figure of speech: A Festschrift for John Laver* (pp. 63–91). Mahwah, NJ: Erlbaum.

Cutler, A., & Norris, D. G. (1988). The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception and Performance*, *14*, 113–121.

Cutler, A., & Otake, T. (1994). Mora or phoneme: Further evidence for language-specific listening. *Journal of Memory and Language*, *33*, 824–844.

de Jong, K. J. (1995). The supraglottal articulation of prominence in English: Linguistic stress as localized hyperarticulation. *Journal of the Acoustical Society of America*, *97*, 491–504.

de Jong, K. J. (2004). Stress, lexical focus, and segmental focus in English: Patterns of variation in vowel duration. *Journal of Phonetics*, *32*, 493–516.

de Jong, K. J., & Zawaydeh, B. A. (2002). Comparing stress, lexical focus, and segmental focus: Patterns of variation in Arabic vowel duration. *Journal of Phonetics*, *30*, 53–75.

Delgutte, B. (1982). Some correlates of phonetic distinctions at the level of the auditory nerve. In: R. Carlson, & B. Granstrom (Eds.), *The representation of speech in the peripheral auditory system* (pp. 131–149). Amsterdam: Elsevier.

Delgutte, B., & Kiang, N. Y. S. (1984). Speech coding in the auditory nerve, I: Vowel-like sounds. *Journal of the Acoustical Society of America.*, *75*, 866–878.

Flege, J. (1995). Second-language speech learning: Theory, findings, and problems. In: W. Strange (Ed.), *Speech perception and linguistic experience: Theoretical and methodological issues* (pp. 233–272). Baltimore, MD: York.

Fougeron, C., & Keating, P. A. (1997). Articulatory strengthening at edges of prosodic domains. *Journal of the Acoustical Society of America*, *101*, 3728–3740.

Keating, P. A. (1984). Phonetic and phonological representation of stop consonant voicing. *Language*, *60*, 286–319.

Keating, P., Cho, T., Fougeron, C., & Hsu, C. (2003). Domain-initial strengthening in four languages. In: J. Local, R. Ogden, & R. Temple (Eds.), *Papers in laboratory phonology VI* (pp. 145–163). Cambridge: Cambridge University Press.

Keating, P. A., Mikos, M. J., & Ganong, W. F. (1981). A cross-language study of range of voice onset time in the perception of initial stop voicing. *Journal of the Acoustical Society of America*, *70*, 1261–1271.

Kim, S., Broersma, M., & Cho, T. (2012). The use of prosodic cues in learning new words in an unfamiliar language. *Studies in Second LanguageAcquisition*, 34 (3).

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.

Kim, S., & Cho, T. (2010). The effect of acoustic correlates of domain-initial strengthening in lexical segmentation of English by native Korean listeners. *Phonetics and Speech Sciences (The Journal of Korean Phonetics Association)*, *2*, 115–124.

Kingston, J., & Diehl, R. L. (1994). Phonetic knowledge. *Language*, *70*, 419–454.

Klatt, D. H. (1975). Vowel lengthening is syntactically determined in connected discourse. *Journal of Phonetics*, *3*, 129–140.

Lehiste, I. (1970). *Suprasegmentals*. Cambridge, MA: MIT.

Pierrehumbert, J., & Talkin, D. (1992). Lenition of /h/ and glottal stop. In: G. Docherty, & D. R. Ladd (Eds.), *Papers in laboratory phonology II: Gesture, segment, prosody* (pp. 90–117). Cambridge: Cambridge University Press.

Quené, H. (1993). Segment durations and accent as cues to word segmentation in Dutch. *Journal of the Acoustical Society of America*, *94*, 2027–2035.

Repp, B. H. (1979). Relative amplitude of aspiration noise as a voicing cue for syllable-initial stop consonants. *Language and Speech*, *22*, 173–189.

Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996). Word segmentation: The role of distributional cues. *Journal of Memory and Language*, *35*, 606–621.

Salverda, A. P., Dahan, D., & McQueen, J. M. (2003). The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension. *Cognition*, *90*, 51–89.

Silbert, N., & de Jong, K. (2008). Focus, prosodic context, and phonological feature specification: Patterns of variation in fricative production. *Journal of the Acoustical Society of America*, *123*, 2769–2779.

Slis, I. H., & Cohen, A. (1969). On the complex regulating the voiced–voiceless distinction I. *Language and Speech*, *12*, 80–102.

Sluijter, A. M. C., & van Heuven, V. J. (1996). Acoustic correlates of linguistic stress and accent in Dutch and American English. In *Proceedings of the fourth international conference on spoken language (ICSLP 1996)* (Vol. 2, pp. 630–633). Philadelphia, PA. USA.

Summerfield, A. Q., & Haggard, M. P. (1974). Perceptual processing of multiple cues and contexts: Effects of following vowel upon stop consonant voicing. *Journal of Phonetics*, *2*, 279–294.

Tyler, M. D., & Cutler, A. (2009). Cross-language differences in cue use for speech segmentation. *Journal of the Acoustical Society of America*, *126*, 367–376.

Van Alphen, P. M., & McQueen, J. M. (2006). The effect of voice onset time differences on lexical access in Dutch. *Journal of Experimental Psychology: Human Perception and Performance*, *32*, 178–196.

Van Alphen, P. M., & Smits, R. (2004). Acoustical and perceptual analysis of the voicing distinction in Dutch initial plosives: The role of prevoicing. *Journal of Phonetics*, *32*, 455–491.

Weber, A., & Cutler, A. (2006). First-language phonotactics in second-language listening. *Journal of the Acoustical Society of America*, *119*, 597–607.

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

Wright, R. (2004). A review of perceptual cues and cue robustness. In: B. Hayes, R. Kirchner, & D. Steriade (Eds.), *Phonetically based phonology* (pp. 34–57). Cambridge: Cambridge University Press.