Studies in Second Language Acquisition

http://journals.cambridge.org/SLA

THE USE OF PROSODIC CUES IN LEARNING NEW WORDS IN AN UNFAMILIAR LANGUAGE

Sahyang Kim, Mirjam Broersma and Taehong Cho

Studies in Second Language Acquisition / Volume 34 / Issue 03 / September 2012, pp 415 - 444
DOI: 10.1017/S0272263112000137, Published online: 15 August 2012

Studies in Second Language Acquisition, 2012, **34**, 415–444.
doi:10.1017/S0272263112000137

THE USE OF PROSODIC CUES IN LEARNING NEW WORDS IN AN UNFAMILIAR LANGUAGE

Sahyang Kim
Hongik University, Seoul, Korea

Mirjam Broersma
Donders Centre for Cognition, Radboud University Nijmegen, and Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands

Taehong Cho
Hanyang University, Seoul, Korea

The artificial language learning paradigm was used to investigate to what extent the use of prosodic features is universally applicable or specifically language driven in learning an unfamiliar language, and how nonnative prosodic patterns can be learned. Listeners of unrelated languages—Dutch (n = 100) and Korean (n = 100)—participated. The words to be learned varied with prosodic cues: no prosody, fundamental frequency (F0) rise in initial and final position, final lengthening, and final lengthening plus F0 rise. Both listener groups performed well above chance level with the final lengthening cue, confirming its crosslinguistic use. As for final F0 rise, however, Dutch

We thank the audience at the meeting of the comprehension group at the Max Planck Institute for Psycholinguistics, especially Anne Cutler and James McQueen, for their valuable comments. We also thank the research assistants for help running the experiment, and all Dutch and Korean participants at Radboud University Nijmegen and Hanyang University in Seoul. This work was supported by the National Research Foundation of Korea grant funded by the Korean government (NRF-2009-332-A00076), given to the first author, and a Veni grant from the Netherlands Organization for Scientific Research (NWO) given to the second author.

Address correspondence to: Taehong Cho, Hanyang Phonetics and Psycholinguistics Lab, Department of English Language and Literature, Hanyang University, Haengdang-dong 17, Seongdong-gu, Seoul (133-791), Korea; e-mail: tcho@hanyang.ac.kr.

listeners did not use it until the second exposure session, whereas Korean listeners used it at initial exposure. Neither group used initial F0 rise. On the basis of these results, F0 and durational cues appear to be universal in the sense that they are used across languages for their universally applicable auditory-perceptual saliency, but how they are used is language specific and constrains the use of available prosodic cues in processing a nonnative language. A discussion on how these findings bear on theories of second language (L2) speech perception and learning is provided.

To understand speech in a second language (L2), listeners have to overcome several challenges. They have to recognize individual speech sounds and locate word boundaries to recognize word forms, retrieve word meanings, and integrate those into a coherent message. Each of those tasks can lead to difficulties that are all but trivial. The current study addresses the challenge of finding word boundaries in running L2 speech, especially in connection with the use of high-level prosodic cues in learning an unfamiliar language.

It is now well known that perception and learning of L2 speech sounds are affected by the phonological properties of listeners' native language (L1) (see Bohn & Munro, 2007; Strange, 1995). For example, native Japanese listeners experience difficulties in perceiving English /r/ and /l/ sounds (Goto, 1971), and American English listeners find it hard to distinguish Hindi dental stops from retroflex stops (Polka, 1991), given that those contrasts do not exist in their L1. Listeners face difficulties even in distinguishing familiar segmental contrasts when they occur in unfamiliar positions (Broersma, 2005; Flege & Wang, 1989), and in perceiving phonetically rich segments in phonologically illegitimate positions (Cho & McQueen, 2006). The phonological influence of the L1 can be so strong that learning a L2—even during childhood—does not guarantee a nativelike perception of some L2 segmental contrasts (Flege & MacKay, 2004). Despite such L1 biases, listeners can learn some, if not all, nonnative sounds via training (Aoyama, Flege, Guion, Akahane-Yamada, & Yamada, 2004; Lively, Logan, & Pisoni, 1993; Logan, Lively, & Pisoni, 1991). Theories of nonnative speech perception that focus on the learnability of individual segments, like the speech learning model (SLM; Flege, 1991, 1995) and the perceptual assimilation model (PAM; Best, 1994, 1995; Best, McRoberts, & Goodell, 2001), claim that not all nonnative sound distinctions are equally difficult for L2 listeners, but that phonetic (dis)similarities between L1 and L2 phoneme categories affect the way L2 learners learn new sound distinctions.

In addition to correctly perceiving and learning individual L2 segments, listeners also have to learn to correctly figure out the boundaries of words and phrases from the continuous speech stream in the L2 (see McQueen, 2005). Studies on lexical segmentation have shown that listeners exploit various cues (e.g., coarticulation, phonotactic, and rhythmic cues) in lexical segmentation of their L1 (Cutler, 1994; Mattys, 2004; McQueen, 1998), and that they tend to use their L1 lexical-segmentation strategies when segmenting a L2. For example, German listeners use German phonotactic constraints for the segmentation of English speech (Weber, 2001), and Japanese listeners have been found to rely on the rhythmic unit of their L1 (the mora) when they process English speech online (Cutler & Otake, 1994). Although these examples clearly suggest that the L1 system affects the processing of nonnative language, little is known about how the native use of segmentation cues influences the use of nonnative segmentation cues, and how those unfamiliar cues in the L2 can be learned and exploited in L2 lexical segmentation.

Another aspect that has been less studied is the role of prosody in lexical segmentation of nonnative speech. Spoken utterances are prosodic in nature. An utterance is prosodic in that its segmental makeup is superimposed by prosodic (or suprasegmental) features such as fundamental frequency (F0), duration, and intensity. It is also prosodic in that such features are employed to build a prosodic structure of the utterance being spoken (Cho, 2002; Keating, 2006; Shattuck-Hufnagel & Turk, 1996). Prosodic structure manifested as such is considered to serve dual functions in speech production—that is, prosodic boundary marking, by which hierarchical grouping of prosodic constituents is determined, and prominence marking, by which relative prominence among prosodic constituents is determined (e.g., Keating & Shattuck-Hufnagel, 2003). The prosodic boundary is generally marked in duration and F0 dimensions as reflected in phrase-final lengthening (Klatt, 1975; Wightman, Shattuck-Hufnagel, Ostendorf, & Price, 1992) and phrase boundary tones (Beckman & Pierrehumbert, 1986; Pierrehumbert, 1980). The duration and F0 dimensions, along with the intensity dimension, are also the acoustic dimensions in which prosodic prominence (i.e., stress and accent) is generally realized (Lehiste, 1970; but see Cho & Keating, 2009, for discussion of how boundary and prosodic markings are phonetically differentiated). A general consensus now is that speech comprehension cannot be understood without understanding the role of prosodic cues that carry the information of high-level prosodic structure.

The use of high-level prosody in lexical segmentation has gained considerable attention since Christophe, Peperkamp, Pallier, Block, and Mehler (2004) showed that listeners terminate a pending lexical search when they encounter a prosodic phrase boundary, and that prosodic structure thus modulates lexical segmentation. Subsequent studies have shown that listeners indeed utilize high-level prosodic cues (i.e., the cues that mark prosodic structure) to segment words from their L1 speech stream (Cho, McQueen, & Cox, 2007; Kim & Cho, 2009; Shatzman & McQueen, 2006; Warner, Otake, & Arai, 2010; Welby, 2007). However, how listeners use high-level prosodic structural aspects of their L1 in lexical segmentation of an unfamiliar language or a L2 has not been fully explored. In this regard, an important issue is the extent to which the use of prosodic cues in nonnative lexical segmentation is universally applicable or specific-language driven.

In most languages, prosodic structure is expressed by a limited set of suprasegmental features, such as F0, duration, and intensity. However, some segmental features such as tongue position, vowel formants, and vowel-to-vowel coarticulation can also be used (Cho, 2004; Cho & Keating, 2009; de Jong, 1995). Given the limited set of acoustic-phonetic features that are available for marking prosodic structure crosslinguistically, some researchers have assumed that the use of prosodic cues for segmentation relies on the universal mechanism of attending to perceptually salient acoustic attributes of prosodic structure. Thus, the acoustic changes in the realm of suprasegmental prosodic features are auditory perceptually salient, such that listeners would attend more to sound units that have longer duration, higher pitch (F0), and stronger intensity than those that do not. For example, Warner, Otake, and Arai (2010) argued that the crosslinguistic use of rising F0 found in Japanese and French (Welby, 2007) as well as Korean (Kim, 2004) is due to the universally applicable auditory-perceptual salience of F0 rise. Phrase-final lengthening has also been claimed to be a universally applicable segmentation cue, as English, Dutch, and French listeners use the final lengthening cue in lexical segmentation (Tyler & Cutler, 2009).

On the contrary, there is also ample evidence for the language-specific use of prosodic cues in lexical segmentation. Kim (2004) showed in an artificial language (AL) learning study that Korean listeners do not make use of a prosodic boundary cue if it does not conform to Korean prosodic phonology. Tyler and Cutler's (2009) crosslinguistic study also showed that Dutch, English, and French listeners use the F0 cue differently, which suggested language-specific use, counter to the findings on the universal use of final lengthening. Kim and Cho (2009) also demonstrated that although the auditory-perceptual saliency of F0 rise may be universal, how such a tonal element is aligned with segmental content is guided by language-specific intonational phonology. In fact, prosodic structure itself is assumed to be language specific (see Beckman, 1996), and so is the way it is phonetically realized (Cho & Jun, 2000; Cho & McQueen, 2005; Ladd, Schepman, White, Quarmby, & Stackhouse, 2009).[1] Just as the detailed manifestation of the acoustic correlates of prosodic structure varies from language to language, so should listeners' sensitivity to and their use of these acoustic cues in segmentation.

The first goal of the current study is therefore to test the universal versus language-specific use of high-level prosodic cues in lexical segmentation by examining how the same set of stimuli, which contain potentially universal F0 and durational cues, are processed by listeners with different L1 backgrounds (i.e., Dutch and Korean listeners). The Dutch and Korean languages were chosen because they are typologically unrelated and prosodically different. If listeners benefited from a potentially universal prosodic cue, even if the cue is not employed by the listeners' L1, the universal view of prosodic cues would be borne out. In contrast, if listeners' segmentation performances were not improved by the presence of a prosodic cue that is potentially universal but nonfunctional in their L1, it would confirm the view that the use of prosodic cues, no matter how auditory-perceptually salient such cues may be, is modulated by language specificity.

The second goal of this study is to examine the extent to which nonnative prosodic structural cues are learnable through extended exposure to the cues. This learnability issue is also related to the question of universal versus language-specific use of prosodic cues. If the use of high-level prosodic cues were universal, listeners would not have to learn new prosodic patterns at all, which would make it relatively easy to learn the new language. In contrast, if the use of these high-level prosodic cues were not universal, listeners would have to depart from their L1 segmentation strategies and refine their perceptual and processing system according to the L2 system, as they do when they learn to distinguish and categorize L2 phonemes. Therefore, this study investigated whether the learnability of a nonnative prosodic structural cue would be further constrained by the listener's linguistic background, like in L2 phoneme learning, and, if so, how the differential learnability could be accounted for in terms of the universality versus language specificity of the cue.[2]

The AL learning technique was employed to examine these issues (see Saffran, Newport, & Aslin, 1996, for more explanations on the AL learning paradigm). In an AL learning experiment, participants are exposed to a relatively long (e.g., 20 min) speech stream of an AL and then tested on whether they have learned novel words that constitute the AL. This experimental paradigm is particularly suitable for the purpose of this study, given that it allows for the manipulation of each prosodic cue at a time, so that the effect of each cue can be seen without any confounding factors (Bagou, Fougeron, & Frauenfelder, 2002; Kim, 2004; Tyler & Cutler, 2009; Vroomen, Tuomainen, & de Gelder, 1998). Dutch and Korean listeners were tested because both languages provide a good test case for the issue of the universal versus language-specific use of prosodic cues—that is, they both employ F0 and duration features in marking prosodic structure but differ substantially in the way the prosodic features are used (see Kim, Cho, & McQueen, 2012, for another case with Dutch and Korean listeners showing that phonetic richness reflected in VOT can outweigh prosodically driven phonological knowledge when learning an AL). Furthermore, to address the learnability issue more in depth, there were two sessions of AL learning. Participants came back for a second experimental session three days after the first session and were tested on their improvement after further exposure to materials with the same prosodic cue.

The use of prosodic cues that have a different status in each language was assessed (see the Experiment: Session 1 section for detailed discussion and relevant references on the different phonological and phonetic statuses of the prosodic cues in Korean and Dutch). First investigated was the use of F0 rise with a high tone in word-final position, which is a phonologically fixed pattern in Korean but not in Dutch. If listeners used this cue in a language-specific way, Korean but not Dutch listeners should rely on it for speech segmentation. The second prosodic cue assessed (i.e., F0 rise with a high tone in word-initial position) is not a systematic prosodic cue in Korean, and, in Dutch, it is not a phonologically fixed pattern but one that occurs frequently in association with initial stressed syllables—this reflects the fact that one of the purposes of the study was to test whether such a common but not fixed pattern would be used in the segmentation of speech in a different language. The third cue—final lengthening—is a common pattern in both Dutch and Korean, and listeners of both languages should benefit from it for speech segmentation. The fourth cue, a combination of final lengthening and final F0 rise, was included to investigate whether their combination would facilitate segmentation to a greater degree than the separate cues—particularly for Korean listeners, who are expected to benefit from both cues, which are systematically used in their L1. Finally, the study tested how the use of these prosodic cues (especially those that are not systematically used in the listeners' L1) can be learned and exploited through repeated exposure, and how their use is related to the universal and language-specific aspects of the cue in question.

The combined results of the two sessions will provide better insights into the mechanism underlying the use of prosodic cues in both L1 and nonnative speech processing. The final goal of this study is therefore to discuss the broader implications of these results in theories of nonnative speech learning (e.g., PAM, Best, 1994, 1995; SLM, Flege, 1991, 1995), and to propose a direction in which existing models of nonnative speech perception should be developed, taking into account high-level prosodic aspects and their role in L2 lexical segmentation.

EXPERIMENT: SESSION 1

The first session of the experiment was designed to investigate Dutch and Korean listeners' exploitation of four different prosodic cues. The Tested Prosodic Cues section discusses how each prosodic cue may be used in Dutch and Korean and what predictions may be borne out of their language-specific versus universal aspects.

Tested Prosodic Cues

F0 Rise with a H Tone in Word-Initial Position (Initial F0 Rise). F0 rise with a H tone or a L + H tone in word-initial position has been known to be a language-specific prosodic cue in Dutch (Tyler & Cutler, 2009; Vroomen et al., 1998) in which the initial F0 cue in the first syllable may mark the frequent trochaic (strong-weak) stress pattern (Quené, 1993).[3] However, about 50% of trisyllabic words—such as those used in this and many other AL learning studies—are produced with penultimate or final stress (Kager, 1989), and duration and spectral balance are the primary cues that mark stressed syllables, whereas F0 movement is secondary (Sluijter & van Heuven, 1996). Moreover, an initial sharp F0 rise with H in the first syllable can also occur initially as a high boundary tone phrase, but it is equally possible that a low boundary tone occurs in that position (Gussenhoven, 2004). Fundamental frequency rise with H in the first syllable in Dutch is therefore not an absolute prosodic cue but one of many prosodic possibilities, which are sometimes probabilistically determined in the given prosodic structure in the sense that only about 50% of trisyllabic words are initially stressed.[4]

In standard Seoul Korean, in contrast, unless the initial consonants are tensed or aspirated stops[5] (which was never the case in the present study), the first syllable of a prosodic phrase generally starts with a L tone as phonologically defined by a LHLH tonal pattern of the accentual phrase (AP), which is an intermediate-level phrase (Jun, 1993, 1998, 2000).[6] The LHLH tonal pattern is used when an AP has more than four syllables, but when an AP consists of less than four syllables, a LH tonal pattern is used with L aligned with the first syllable and H with the last syllable. Therefore, the initial F0 rise with H in the first syllable is not seen as an optimal prosodic cue, at either the word or the phrasal level, except when the initial segments are tensed or aspirated stops (Jun, 2000; Kim, 2004; Kim & Cho, 2009). Given the crosslinguistic difference in the use of initial F0 rise, if listeners show language-specific modulation of prosodic cues in lexical segmentation, it is possible to predict that Korean listeners will not benefit from the cue—although it remains to be seen whether Dutch listeners will use this cue. However, if both listener groups benefit from the initial F0 rise, the universal use of F0 rise will be supported, in line with Warner et al. (2010).

F0 Rise with a H Tone in Word-Final Position (Final F0 Rise). Unlike F0 rise in word-initial position, both languages employ F0 rise in word-final position to mark the end of a prosodic phrase, but its phonological status is different in the two languages. As introduced previously, F0 rise with H in the final syllable is a phonologically fixed pattern at the end of the Korean AP. In Dutch, however, the phrase-final F0 rise (i.e., high boundary tone) is simply one of the possible intonational patterns (Gussenhoven, 2004). The languages also differ in that the Korean F0 rise at the end of the AP is used independently, seldom accompanied by phrase-final lengthening (Jun, 1993, 1998, 2000; but see Cho & Keating, 2001), whereas in Dutch the phrase-final F0 rise with H in the final syllable usually comes with substantial phrase-final lengthening (Cambier-Langeveld, 2000; Gussenhoven, 2004).

Given that lexical segmentation is facilitated when the word-final position is matched with the phrase-final position—for example, AP-final position, even in the middle of the utterance, which might comprise several APs (Cho et al., 2007; Christophe et al., 2004)—and also given that both languages use final F0 rise in marking a phrase-final position, F0 rise would give listeners the percept of phrase finality and thus help them to locate a possible word boundary from a continuous speech string, as previous studies have independently shown (Kim, 2004, for Korean listeners and Tyler & Cutler, 2009, for Dutch listeners).[7] However, the apparent crosslinguistic difference in the phonological use of final F0 rise may yield differential degrees of benefit: Korean listeners might make more out of the recurring final F0 rise than Dutch listeners, as it is a more consistent phonological marker of the phrase-final (therefore, word-final) position in Korean. After all, final F0 rise without phrase-final lengthening may not be a robust lexical segmentation cue for Dutch listeners because—unlike in Korean—final F0 rise in Dutch always comes with phrase-final lengthening. If final F0 rise is exploited simply as a universal prosodic cue, however, the degree of benefit should be comparable for listeners of both languages.

Final Lengthening. Phrase-final lengthening was included because it is a potentially universal auditory-perceptual cue and a common feature that exists in both Dutch and Korean as a prosodic marker of the end of the intonational phrase (the largest prosodic phrase often assumed across languages, see Jun, 2004). It is thus expected that listeners of both languages will benefit from the final lengthening cue, regardless of whether the cue is universal or language specific.

Final Lengthening and F0 Rise. Finally, the final lengthening cue was combined with F0 rise to assess whether a combination of potential prosodic cues to phrase-final position can help lexical segmentation in a cumulative way—that is, whether listeners benefit more when they are exposed to lengthening and F0 rise cues together, as opposed to when they are exposed to one cue at a time. Some previous work has suggested that listeners might be able to use available segmentation cues in a cumulative way (Spitzer, Liss, & Mattys, 2007). If this were the case, both Korean and Dutch listeners' learning performance may improve with the use of combined cues as opposed to just one prosodic cue. However, in an AL learning study with French listeners, the combination of final duration and pitch cues was not found to be significantly more helpful than a final duration cue alone (Bagou et al., 2002). In a similar manner, Kim and Cho (2009) did not find any cumulative effect of final lengthening and final F0 rise in a word-spotting task with Korean listeners. Kim and Cho argued that final lengthening is a primary cue that listeners rely on in lexical segmentation, and that the pitch cue can come into play only if the lengthening cue is not fully available. Therefore, it remains to be seen in the present study whether or not a cumulative effect is to be found in AL learning by both Dutch and Korean listeners.

METHOD

Participants

Two hundred participants, 100 native listeners of Dutch and 100 native listeners of Seoul Korean, participated in the study. The Dutch participants were college students at the Radboud University Nijmegen, the Netherlands, whereas the Korean participants were college students at Hanyang University in Seoul, Korea. Both groups of participants were all in their early to mid twenties, had normal hearing, and were paid for their participation.[8]

Materials

An AL with six trisyllabic words was created. Four consonants (/p, t, k, m/) and four vowels (/a, i, u, ɛ/), which exist in both the Dutch and the Korean phoneme inventories, were selected. The eight segments were exhaustively combined to make 16 distinct consonant-vowel (CV) syllables. These syllables were further combined to make six trisyllabic words: [tikɛpu], [pɛtami], [mupaki], [kapimɛ], [kutɛpa], and [pimatu]. None of the six words were identical or similar to real words in Dutch or Korean.

A female L1 speaker of Korean recorded each CV syllable 10 times. Each syllable was produced independently, following a pause, to avoid coarticulation across syllables (for the effects of coarticulation in lexical segmentation, see Johnson & Jusczyk, 2001; Mattys, 2004).[9] It is important to note that the use of a Korean speaker could have affected the listeners' performance, but it turned out that the Dutch and Korean listeners showed exactly the same performance in the baseline (no-prosody) condition (mean percentage of correct answers: 62.7% for Dutch vs. 62.3% for Korean listeners).

For each CV syllable, one token—of the 10 recorded—was selected that was most clearly articulated, with minimum F0 movement within the syllable. For the selected tokens, duration, intensity, and F0 of each individual syllable were normalized to the average value of all the syllables. The normalized syllables were concatenated to form the six trisyllabic words. Finally, 126 repetitions of the six words were concatenated in random order, which yielded a speech stream of approximately 10 or 12 min (the final lengthening as well as the final lengthening and F0 rise conditions lasted 12 min, and all other conditions lasted 10 min). No word occurred twice in a row, and there was no pause between the words in any part of the speech stream. Syllable-to-syllable transitional probability (i.e., the frequency of y following x, given the frequency of x) within words ranged from 0.5 to 1, and the probability across words ranged from 0.03 to 0.44. Saffran and colleagues (1996) showed that higher transitional probabilities within words help listeners to group the sequence as a word, thus facilitating lexical segmentation.

The experiment had five prosodic conditions that applied to the three-syllable target words: (a) no prosody (the baseline condition in which all syllables had the same duration, pitch, and intensity), (b) initial F0 rise (high-pitch cue on the initial syllable), (c) final F0 rise (high-pitch cue on the final syllable), (d) final lengthening (lengthened duration cue on the final syllable), and (e) final lengthening and F0 (both lengthened duration and high-pitch cues on the final syllable).[10] The resulting AL in each condition consisted of target words that had the exact same prosodic patterns. Figure 1 illustrates each condition with wave forms and corresponding F0 contours.

Digital speech-analysis software (Praat) was used for the manipulation and concatenation of the materials. F0 rise was made by increasing the baseline tone from 190 Hz to 250 Hz, equivalent to 4.75 semitones. It is important to note that Tyler and Cutler (2009) used a more extreme F0 difference, increasing F0 by approximately 6 semitones. The current study followed Sluijter and van Heuven (1996), who suggested that the F0 difference should be minimally more than 4 semitones for lexical stress in Dutch, and Lee (1998), who modeled phrase-final F0 rise in Korean by just more than 4 semitones. The final lengthening condition was created by making the final syllable longer than the baseline syllable by 1.77 times (from 252 to 446 ms), taking into account both Korean and Dutch production data (Cambier-Langeveld, Nespor, & van Heuven, 1997; Cho & Keating, 2001; Cho & McQueen, 2005; Chung et al., 1996). Both the pitch and the duration manipulations were made using the overlap-add synthesis function, also known as TD-PSOLA (time-domain pitch-synchronous overlap-and-add), available in Praat, which allowed manipulation of pitch without distorting durational information, and of duration without distorting pitch information.

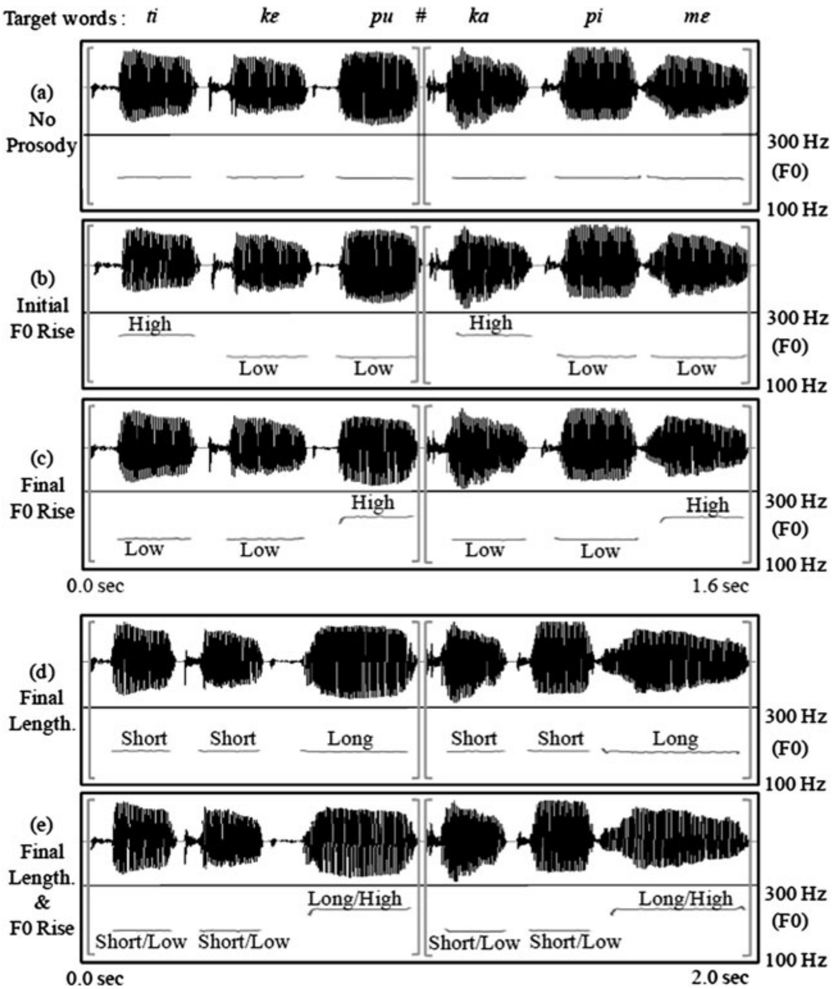

Figure 1. Sample sound waveforms and corresponding F0 contours for the test word sequence [tikεpu]#[kapimε] with five conditions: (a) no prosody, (b) initial F0 rise, (c) final F0 rise, (d) final lengthening and (e) final lengthening and F0 rise. In the actual experiment, listeners heard sequences of all 6 test words in different orders for 20–24 min.

Procedure

Participants were randomly but evenly distributed over the five experimental conditions, and they were tested individually or in groups in a quiet room. The experiment was composed of a learning phase and a testing phase. During the learning phase, the participants heard a speech stream from one of the five conditions. They were told that they would hear a speech stream from a simple AL that was composed of concatenated nonsense words, and that there would be no pause between words. They were informed that their task was to find words of the AL from the speech stream, but they did not receive any information about the number of words in the language. The participants heard the stimuli from a PC through high-quality headphones, and they were asked to adjust the volume to a comfortable level. During the learning session, they heard the approximately 10-to-12-min-long concatenated string, followed by a 1-min silent break (during which no talking was allowed), and then heard the same 10-to-12-min string again. There were 20-ms fade-in and fade-out periods for each learning session, such that participants would not get any information about word boundaries at the beginning or the end of the speech stream.

For the testing phase, 36 forced choice pairs were made from the combination of the six trisyllabic existing words of the artificial language and six trisyllabic foils. The foils were composed of three part-word and three nonword strings, both of which did not exist in the AL. Part-word foils contained the final two syllables of an existing word plus an additional syllable. The transitional probabilities between the overlapping string and the additional syllable were 0.39–0.44, which was less than the transitional probability range of 0.5–1.0 within existing words. Nonword foils were composed of the syllables that were used in the learning phase, but in a sequence that participants had never heard. Thus, the transitional probability within nonwords was zero. In each trial, participants heard a pair of trisyllabic strings, one of which was an existing word and the other of which was a foil, with an 800-ms interstimulus interval. Strings presented in the testing phase had no prosodic cue (as in the no-prosody condition). At the beginning of the test phase, participants were given an answer sheet with the response options 1 and 2, which corresponded to two spoken stimuli, only one of which had been used as a word in the AL. They were instructed to indicate on each trial whether the first or the second stimulus word was the one that they thought existed in the newly learned AL. They had 4 s to indicate their response before the next trial started.

RESULTS

A series of one-sample *t* tests on prosodic conditions showed that Dutch listeners performed at an above-chance level—with a chance level of 0.5—in each of the five conditions ($p < .001$ in all conditions), as shown in the left panel of Figure 2. Korean listeners showed above-chance performance in all conditions ($p < .001$), except the initial F0 rise condition, as shown in the right panel of Figure 2. This indicates that listeners of

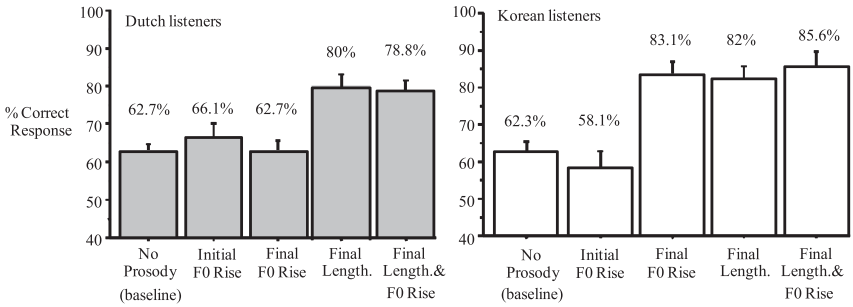

Figure 2. Percentage of correct responses in each prosodic condition in session 1: Dutch listeners (left) and Korean listeners (right).
Note: Error bars indicate standard errors.

both language groups managed to learn the new language using the transitional probabilities (in the no-prosody condition) and the prosodic cues (in all other conditions).

A two-way ANOVA was performed with the factors language (Dutch vs. Korean participants) and prosody (no prosody, initial F0 rise, final F0 rise, final lengthening, and final lengthening and F0 rise). There was a main effect of prosody, $F(4,190) = 16.618$, $p < .001$. The effect of language just missed significance, $F(1,190) = 3.841$, $p = .051$. Crucially, the interaction between language and prosody was significant, $F(4,190) = 4.901$, $p = .001$. With separate ANOVAs for each language group, the effect of prosody was significant for Dutch listeners, $F(4,95) = 9.051$, $p < .001$, as well as Korean listeners, $F(4,95) = 11.755$, $p < .001$.

Bonferroni post hoc tests were done to assess the predictions made for each condition. In the initial F0 rise condition, Korean listeners did not perform better than in the control (no-prosody) condition. In that condition, the Dutch listeners did not perform better than in the control condition either—thus, they did not benefit from the initial F0 rise pattern for segmentation of the new language. In the final F0 rise condition, Korean listeners performed better than in the control condition ($p < .01$), whereas Dutch listeners' performance did not differ from the control condition. As for the final lengthening condition, both Dutch and Korean listeners performed better than in the control condition ($p < .01$). Finally, in the final lengthening and F0 rise condition, both Dutch and Korean listeners performed better than in the control condition ($p < .01$). The combined cue, however, did not give rise to better performance as compared to the final lengthening cue alone for both Dutch and Korean listeners, and as compared to the final F0 rise cue alone for the Korean listeners. Additionally, in the final F0 rise condition, the Korean listeners performed significantly better than the Dutch listeners, $F(1,38) = 20.910$, $p < .001$. Dutch and Korean listeners did not differ significantly in any other condition.

DISCUSSION

The results clearly show that the listeners' L1 did affect segmentation of the newly learned AL. Both Dutch and Korean listeners did not rely on a universal F0 rise cue in word-initial position. Initial F0 rise is only probabilistically determined in Dutch phonology, and the pattern does not exist as a possible prosodic pattern in Korean. Additionally, Korean listeners but not Dutch listeners were found to rely on the final F0 rise cue, as predicted on the basis of the fact that final F0 rise is phonologically defined in Korean, but not in Dutch. Both Dutch and Korean listeners were found to benefit from the final lengthening cue, which corresponds to a prosodic boundary marker in both languages (Cambier-Langeveld, 2000; Gussenhoven, 2004; Jun, 2000). This finding is in line with previous findings with other languages, including English and French (Bagou et al., 2002; Tyler & Cutler, 2009).

It is necessary to point out that no cumulative effect was found with the combined final lengthening and F0 rise cues. Kim and Cho (2009) suggested that the durational cue is the most robust segmentation cue, and that the presence of an additional F0 cue is helpful only when the final lengthening cue is not fully manifested in the speech signal. On the basis of the current findings, it is not possible to determine whether the final lengthening cue is more robust than the final F0 rise cue for Korean listeners—it certainly is for Dutch listeners, who did not use the latter cue—but the results confirm that there is no additive effect of the two cues above the effect of the individual cues.

For the Korean listeners, a F0 manipulation led to a significant improvement compared to the control condition when it occurred in final position, but no improvement was observed when a F0 rise of the exact same size occurred in initial position. The Korean listeners' effect of final F0 rise was comparable to that in the final lengthening as well as the combined final lengthening and F0 rise condition. The Korean listeners' high score in the final F0 rise condition (i.e., 83.1% correct) suggests that Korean listeners fully exploited the final F0 rise cue. The lack of an effect of F0 rise in initial position and the full use of the same F0 rise in final position clearly support the view that proposes language specificity in use of prosodic cues. As Kim and Cho (2009) postulated, F0 rise may be universal in terms of its auditory-perceptual saliency, but how it is aligned with segment content falls within the realm of language-specific intonational phonology. More specifically, F0 rise is more likely to be associated with phrase finality in Korean than in Dutch, such that Korean listeners appear to make more use of the F0 rise cue in finding a word boundary in processing nonnative language.

To test whether prosodic cues that are not initially exploited in AL learning might be exploited later when learners receive more exposure to an AL with such cues (i.e., can listeners learn to use prosodic cues through mere exposure, given enough opportunity to learn those cues?), a second experiment with the same participants was carried out in a second session of AL learning. The results are presented in the Experiment: Session 2 section.

EXPERIMENT: SESSION 2

Listeners came back for a second session, to test whether cues that had not fully been exploited in the first session (i.e., initial F0 rise and final F0 rise for Dutch listeners, and initial F0 rise for Korean listeners) could be learned and exploited when listeners were repeatedly exposed to the same prosodic cues in the second session. In the case of the Dutch listeners, all 100 participants were tested twice. However, the data for the Korean listeners from the first session were almost simultaneously analyzed, and it was possible to monitor which cues were not exploited in the first session (the Dutch data from the first session could not be simultaneously analyzed). Therefore, of the Korean listeners, only the 40 participants in the no-prosody and initial F0 rise conditions—which both showed a null effect—were brought in for a second session.

For each L1 and condition, participants who had taken part in a particular condition were randomly divided into two groups of 10. The members of the first group were exposed to the exact same AL that they had heard in session 1 (henceforth the *same* group). The members of the second group were exposed to the same prosodic condition that they had heard in session 1, but to new lexical items with different syllable sequences (henceforth the *different* group). For this group, a new AL was constructed based on the same speech materials used for session 1.

The different condition was included to exclude the possibility that listeners might perform better in session 2 than in session 1 simply because they remembered the words of the AL from the previous session, rather than because of an improved processing of the prosodic structure of the AL. Given that the elements that are supposed to be learned, in addition to transitional probabilities of the syllable sequences, are suprasegmental prosodic features, if the different group showed improved performance, it would allow generalization that the learning effect is purely prosodic—that is, independent of segmental makeup.

In both groups, participants thus listened to the same prosodic pattern they had been exposed to in session 1. For example, a participant who had been exposed to the initial F0 rise condition in session 1 would listen to the same initial F0 rise condition in session 2 as well, regardless of whether he or she belonged to the same or the different group.

METHOD

Participants and Procedure

Participants from session 1 came back 3 days after that session to take part in the second session. All of the 100 Dutch participants and 40 of the Korean participants—only those who had participated in the initial F0 rise and the no-prosody conditions—took part. Half of the listeners (10 per condition) had the same learning and testing materials as in session 1 (the same group), and the other half (10 per condition) were exposed to a new AL, with the same prosodic conditions as in session 1 (the different group). The procedure was exactly the same as in session 1.

Materials

For the different AL, a new set of words was created. There were six trisyllabic words, as in session 1, and the prosodic patterns were also the same as in session 1. The six words were [putaki], [makutɛ], [tipɛmu], [kapitu], [mikɛpa], and [kumɛpi]. Transitional probabilities were again 0.5–1 within words and 0.03–0.44 across words. The testing phase was similar to that in session 1, but included the six different words and newly constructed part-word and nonword foils. In the same language, the materials were exactly the same as those in session 1; however, during the test phase, the test items were presented in a different order than in session 1.

RESULTS

Dutch Listeners

A series of one-sample *t* tests revealed that Dutch listeners performed above chance (with a chance level of 0.5) in all conditions ($p < .05$), except for the no-prosody condition of the same AL. There was a significant main effect of prosody for the Dutch listeners, $F(4,95) = 6.431$, $p < .001$, but no effect of AL (same vs. different), nor was there an interaction between prosody and AL. The results from the same and the different languages were combined for further statistical analyses, and Bonferroni post hoc analyses showed that, as in session 1, Dutch listeners were significantly better in the final lengthening as well as final lengthening and F0 rise conditions than in the control (no-prosody) condition ($p < .05$ and $p < .001$, respectively).

The results of session 2 were further compared with those of session 1, to examine whether listeners showed a learning effect from repeated exposure, and whether their experience with the prosodic cue in the first session was carried over to a novel set of words in the different language condition. A repeated measures ANOVA was performed with session (session 1 vs. session 2) and prosody as within-subject factors and with AL (same vs. different) as a between-subject factor. There was a significant main effect of session, $F(1,18) = 5.261$, $p < .05$: Listeners' percentage of correct responses improved significantly in session 2 (mean 73.1%), compared to their performance in session 1 (mean 70.1%). There was also a main effect of prosody, $F(4,72) = 8.952$, $p < .001$, and a trend toward a Session × Prosody interaction, $F(4,72) = 2.249$, $p = .072$.

To examine the detailed progress from session 1 to session 2, separate repeated measures ANOVAs were performed with the factors session (session 1 vs. session 2) and prosody with only two conditions at a time (the baseline no-prosody condition vs. one of the four prosodic conditions). As shown in Figure 3, listeners' performance was generally better in session 2 than session 1 in the conditions that involved F0 (i.e., initial F0 rise, final F0 rise, and final lengthening and F0), but a significant difference between the two sessions was found only with the final F0 rise condition. In the final F0 rise condition, Dutch listeners' percentage of correct responses in session 2 (mean 71.2%) was significantly higher than that in session 1 (mean 62.7%; $F(1,18) = 5.533$, $p < .05$). In that condition, there was also an interaction between session and prosody, $F(1,18) = 7.669$, $p < .05$, which suggested that listeners were better with final F0 when compared to the baseline condition (i.e., no prosody).

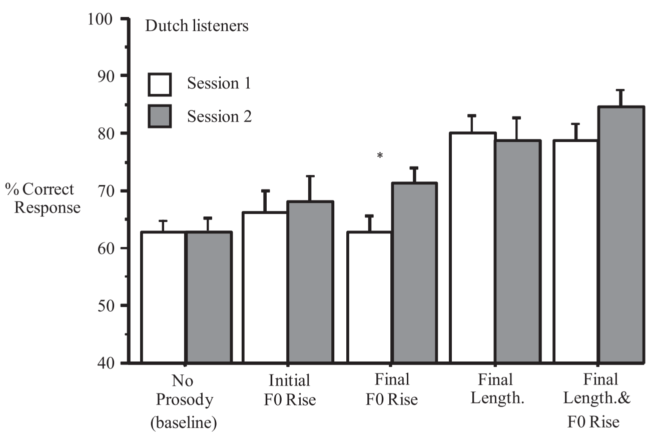

Figure 3. Percentage of correct responses in each prosodic condition: session 1 (white bars) and session 2 (gray bars) for Dutch listeners. *Note*: * marks a significant difference between the two sessions at $p < .05$. Error bars indicate standard errors.

Korean Listeners

A series of one-sample *t* tests revealed that Korean listeners performed above chance (with a chance level of 0.5) in both conditions ($p < .05$), with both the same and the different AL. A repeated measures ANOVA was performed with the two within-subject factors session (session 1 vs. session 2) and prosody (no prosody vs. initial F0 rise), and with the between-subject factor AL (same vs. different). None of the three factors yielded a significant effect, nor was there a significant interaction between session and prosody—that is, there was no improved performance with initial F0 in session 2, as illustrated in Figure 4.

DISCUSSION

The significant main effect of session indicates that Dutch listeners generally performed better in session 2 than in session 1. This overall improvement, however, was largely due to the conditions involving F0 cues, especially with F0 rise in final position, which was the only condition that gave rise to statistically significant improvement. The final lengthening as well as the final lengthening and F0 rise conditions—which had already shown very high correct percentages—did not show any further improvement, presumably due to a ceiling effect.

One of the most important findings was that Dutch listeners' performance with the final F0 rise improved significantly in session 2. This suggests that during their second exposure to the F0 rise cue in final position, listeners somehow came to realize that the cue carried useful boundary information (e.g., marking the end of words), and so they started to attend to the cue for lexical segmentation.

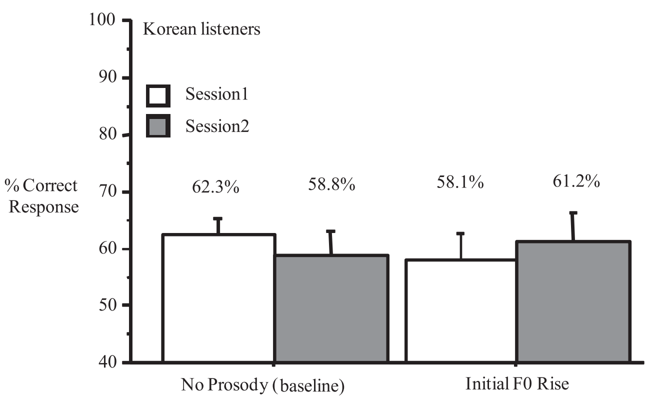

Figure 4. Percentage of correct responses in each prosodic condition: session 1 (white bars) and session 2 (gray bars) for Korean listeners. *Note*: Error bars indicate standard errors.

As discussed in the introduction, the final F0 cue by itself is not a phonologically determined cue in Dutch, but it is a possible cue that can contribute to marking the end of a prosodic unit, which is most frequently combined with phrase-final lengthening. Therefore, Dutch listeners are not familiar with its use without final lengthening. The results of session 2 suggest that listeners might not actively exploit prosodic cues that are not phonologically determined prosodic markers in the L1 when they are exposed to the cue initially, but listeners can learn to use such cues when they are exposed to them for a longer time. In contrast with the final F0 rise condition, Dutch listeners did not show learning in the initial F0 rise condition, which implied that not all prosodic cues used in a novel language can be learned through simple extended exposure in the two experimental sessions.

The asymmetry between initial and final F0 rise conditions indicates that although both initial F0 rise and final F0 rise were not optimal cues for Dutch listeners, F0 rise was a better indicator of final than initial position (at least for trisyllabic words). The differential use of F0 rise depending on position (initial vs. final) appears to have to do with the way these cues are used in the L1. In Dutch, F0 rise is used more systematically in the final position than in the initial position. Final F0 rise most frequently occurs with the most robust cue, the phrase-final lengthening, whereas initial F0 rise is probabilistically linked with the trochaic stress pattern. This is especially true for monosyllabic and disyllabic words, but not necessarily for trisyllabic words (Kager, 1989)—such as the words used as lexical items in the present study. Although neither of them is phonologically determined in Dutch, the relatively more stable use of F0 rise in final position in Dutch appears to account for Dutch listeners' learning to use this cue for the AL segmentation. Korean listeners did not show any learning of F0 rise in initial position either, which can be accounted for more straightforwardly: When a prosodic cue does not have any phonetic or phonological correspondence in the L1 (F0 rise in initial position is not even possible in Korean), the cue is not learnable through simple repeated exposure of the type provided in this experiment.

Another important finding was that there was no statistical difference in the percentages of correct responses of listeners who were exposed to the same AL and in the percentages correct of those exposed to the different AL (with different lexical items). These results suggest that the learning effect found for the Dutch listeners in session 2 compared to session 1 was not attributable to the memorization of the words (i.e., syllable sequences) that they had heard in session 1. As for the Dutch listeners' improved performance with final F0 rise, this implies that the prosodic knowledge that was obtained through the first exposure to the AL was carried over, and that what was learned was purely prosodic.

GENERAL DISCUSSION

The goal of this study was to explore the extent to which prosodic features can be taken as universal or as language specific in terms of their use in nonnative lexical segmentation, and the extent to which nonnative prosodic patterns can be learned. To explore these issues, an AL learning paradigm was used, in which listeners of two typologically unrelated languages, Dutch and Korean, participated. One of the most important findings of the present study was that Dutch and Korean listeners learned the same AL (composed of the same physical speech materials and lexical items) sometimes in much the same way and sometimes in a very different way. This reflected both the universal and the language-specific ways of learning an unfamiliar AL. The universality versus language specificity was determined by which prosodic cues were used and how they were used. For example, a similar crosslinguistic pattern of AL learning was observed for Dutch and Korean listeners when the available prosodic cue in the speech input was a durational cue (lengthening) in word-final position (either by itself or with additional F0 rise), which was consistent with previous studies on different languages such as English and French (Bagou et al., 2002; Kim, 2004; Tyler & Cutler, 2009). The use of the durational cue in speech comprehension can be said to be universal in the sense that it is auditory-perceptually salient, it is used to signal the end of a prosodic phrase, and it is used as a lexical segmentation cue crosslinguistically.[11]

The use of the F0 cue, too, may be universal, in the sense that it has a universally applicable perceptual saliency (Warner et al., 2010), but it is language specific in the way the tonal element is aligned with segmental content (Kim & Cho, 2009). In this study, when the word-final position was associated with F0 rise, Korean listeners' performance in AL learning was greatly facilitated at initial exposure after a single training session (as compared to the baseline, or no-prosody condition), but Dutch listeners did not show such facilitation until after the second training session. This crosslinguistic difference with final F0 rise in processing the nonnative AL can be understood as coming from the final F0 cue's phonological status. In Korean the use of F0 rise in final position is phonologically determined to be an obligatory tonal pattern, marking the end of a prosodic phrase. This phonologically categorical way of using the final F0 cue leads to its full use by L1 listeners of Korean in processing an unfamiliar language, resulting in Korean listeners' very good performance in the first session. In Dutch, on the contrary, although the use of F0 rise in final position (the final F0 rise condition) is phonologically possible, it is one of the optional cues for marking a phrase-final position. When such an optional cue exists in the nonnative language, listeners do not seem to utilize it fully at initial exposure—as evident in the null effect in the first session—but they can gradually learn to use it in lexical segmentation as they are exposed to it more (evidenced by their improved performance in the second session).

The language-specific way of using F0 can also be seen in the way that F0 rise is aligned with segments in initial position. The results showed that neither Korean nor Dutch listeners made use of the same physical F0 rise in AL learning when it was associated with the initial syllable of the target word. Their performance did not show any improvement, even when they were exposed to the same AL for the second time. It is necessary to recall that Warner and colleagues' (2010) claim on the universal use of F0 was based on the fact that Japanese listeners benefited from the presence of initial F0 rise in lexical segmentation, which is generally in line with the findings of Welby (2007), who showed that French listeners exploit an early F0 rise in finding a word boundary. However, Japanese and French differ in that the initial F0 rise in Japanese generally spreads into the second mora, whereas the French F0 rise comes earlier, often associated with the first syllable, which shows language-specific ways of aligning F0 rise with segments. In this study, Korean and Dutch listeners found it hard to use the F0 rise cue in the first syllable in learning the AL, again precisely because Korean and Dutch differ from French or Japanese in terms of how and where in the word the auditory-perceptually salient cue is aligned with segmental content.

However, Korean and Dutch are also different from one another in using initial F0. The Korean prosodic phonology does not allow F0 rise in initial position, not even as a possible phonetic pattern (except when the initial segments are tensed or aspirated, which was never the case in the present study). In contrast, F0 rise is a possible but nonsystematic phonetic correlate of word-initial position in Dutch, but its use is not as systematic as F0 rise in final position. In Dutch, final F0 rise is more systematic than initial F0 rise in that final F0 rise is always associated with consistent phrase-final lengthening, whereas the occurrence of initial F0 rise is optional and probabilistically linked with an initial stress pattern. This language-specific difference in initial F0 rise, however, has not been reflected in the AL learning results, as neither Dutch nor Korean listeners benefited from initial F0 rise for learning the AL. The only piece of evidence that might support the different status of the initial F0 rise in Dutch and Korean was that Korean listeners with initial F0 rise showed at-chance-level performance in the first session; this performance was (numerically but not statistically significantly) poorer than in the baseline condition (no prosody). This possible inhibitory effect was presumably due to the interference of the Korean final F0 rise pattern. Thus, the F0 rise in initial position may have been erroneously interpreted as a final F0 cue and may have interfered with learning of the AL. The interference was robust enough to override the benefit that comes from transitional probability cues, which resulted in a learning effect that was not significantly above chance.

The results of Tyler and Cutler's (2009) study also provide evidence for the different status of initial F0 between Dutch and Korean, as they showed that the initial F0 rise could indeed be exploited in AL learning by Dutch listeners when the F0 rise was high enough. It is important to point out that F0 was raised by 4.75 semitone in these stimuli, versus 6.03 semitone in Tyler and Cutler's stimuli. Given that Tyler and Cutler's study used a substantially higher F0 rise than the present study, the discrepancy between the two studies may be interpreted as suggesting that the use of the initial F0 rise is phonetic in the sense that its effect is determined by the numeric F0 difference, and not by the categorical high versus low tone difference.

Whether the combination of final lengthening and final F0 rise would yield a cumulative effect was also tested, and no cumulative effect was found. This suggests that, although there were two universally perceptually salient cues in the input, the prosodic function they serve is the same as the function served by a single cue—for instance, final duration alone. Thus, with or without F0 rise, the final duration cue was enough to mark a prosodic boundary. This indicates that when an individual prosodic cue (e.g., a universally applicable phrase-final lengthening cue) is strong enough to give out prosodic boundary information to the listeners, the presence of an additional cue (e.g., the language-specific final F0 rise) in the same position does not necessarily help them more. This is in line with the results found in an AL learning study on French listeners (Bagou et al., 2002), and in a word-spotting task with Korean listeners (Kim & Cho, 2009). It should be noted, however, that the results could simply be task specific because there are studies that indirectly suggest the possibility of a cumulative effect in the use of segmentation cues (Dilley & McAuley, 2008; Spitzer et al., 2007).

All in all, the results of this study suggest that the use of prosodic cues is universal in the sense that duration, F0, and intensity are used crosslinguistically in marking important landmarks of prosodic structure, such as initial and final positions. It is also universal given the fact that all these prosodic features are auditory-perceptually salient to the ears of listeners crosslinguistically. At the same time, the use of prosodic cues is language specific, in that the exact way in which a prosodic cue is used is modulated by the prosodic phonology of a given language. Even though F0 rise is auditory-perceptually salient and therefore is used crosslinguistically, the way it is used is modulated by the prosodic phonology of a given language. Even the use of the phrase-final lengthening cue in lexical segmentation, which is considered universal (e.g., Tyler & Cutler, 2009), can be seen as a consequence of its language specificity—that is, it may simply be the case that many languages happen to have a similar language-specific way of using the cue. Therefore, it seems the case that it is the language-specific way of using prosodic features, and not a universally driven auditory-perceptual saliency itself, that determines the way listeners process both native and nonnative speech.

The present study also questioned whether nonnative (unfamiliar) prosodic cues that are initially difficult to learn (in the first session) could be learned and exploited in lexical segmentation through repeated exposure to them, and how learnability of nonnative prosodic cues is constrained by the learner's L1. The crosslinguistic results showed that, although there was a general tendency toward improved performance as a result of the second learning session, the only cue that showed a statistically significant improvement in the present study was the final F0 rise by Dutch listeners—Korean listeners did not go through the second session with this condition because they had already shown a very high percentage of correct responses in the first session. The initial F0 rise, in contrast, was still not learned in the second session by either Dutch or Korean listeners, although it has been shown that initial F0 rise can be exploited in AL learning by Dutch listeners (Tyler & Cutler, 2009). These results suggest that adults have an ability to learn a nonnative prosodic cue through passive repeated exposure, and to use the cue in word segmentation, but only when the nonnative prosodic cue exists in the L1 as a possible phonetic correlate of prosodic structure. Furthermore, the fact that the improvement was made even when listeners were exposed to a different AL (composed of different lexical items) suggests that the knowledge that was carried over was purely prosodic, independent of its segmental makeup.

The findings of the present study have implications for theories of nonnative speech perception such as the SLM (Flege, 1991, 1995) and the PAM (Best, 1994, 1995). Based on the phonetic and phonological similarities and differences between L1 and L2 sounds, SLM divides L2 sounds into three categories—namely, identical, similar, and new. The perceptual assimilation model describes four ways in which L2 sounds can be mapped onto L1 sound categories: two-category assimilation, single-category assimilation, category-goodness assimilation, and non-assimilation. Using these categories, each theory predicts the relative degree of difficulty in perceiving and learning the distinctions between nonnative phones. These models, however, have been developed primarily in terms of learnability and perceptibility of individual phonemes and their phonological distinctions at the segmental level, without taking into account the role of high-level prosodic structure in speech comprehension in general and in nonnative listening in particular. The results of this study can be considered in terms of how the existing models of nonnative speech perception can be further developed. On the basis of the observations about the (dis)similarities of the prosodic patterns in the AL with native prosodic patterns, at least three degrees of relative difficulty in learning and utilizing nonnative prosodic cues can be identified, in a similar way as SLM categorizes learning of L2 sounds.

The first is the identical-prosodic pattern. In this case, the prosodic feature exists in both L1 and L2, and its phonetic and phonological statuses are the same in both languages (e.g., final lengthening). Because the same prosodic category with the same phonetic manifestation exists in their L1, L2 learners should not have difficulty in using these prosodic cues in segmentation, even when they encounter the L2 for the first time. This is exactly what was found with both groups of listeners in the final lengthening as well as the final lengthening and F0 rise conditions. The case of Korean listeners in the final F0 rise condition also falls into this category, given that the AL and Korean both employ the cue in the same way. The second category is the similar-prosodic pattern. This is the case when both L1 and L2 have the same tonal category, but its phonological status and phonetic manifestation varies in L1 and L2. An example of this category from this study would be final F0 rise for Dutch listeners. The Dutch language includes a tonal category of a high phrase-final boundary tone, yet the cue is almost always phonetically realized with a concomitant lengthening cue (e.g., Cambier-Langeveld, 2000). A high F0 cue alone, therefore, is not an optimal phonetic manifestation of phrase-final boundary in Dutch. As seen in the experiment, listeners may initially experience difficulty in dealing with such prosodic patterns, but they can learn to use the prosodic pattern with repeated exposure to the cue. The third category—the new prosodic pattern—encompasses prosodic patterns that exist only in the L2. An example from this study would be Korean listeners' encounter with high F0 in initial position, which is neither phonologically permissible nor phonetically possible in that position in Korean. This category will be the most difficult prosodic pattern to learn because listeners have to create a new category that does not exist in their native prosodic structure. Indeed, in this study, Korean listeners did not manage to learn to use this cue for segmentation, even after two training sessions.

It should be noted that the SLM predicts that learning a L2 segmental category that is similar to an existing category in L1 is more difficult than learning a new segmental category because learners tend to substitute the existing L1 segmental category for the similar L2 segment (Flege, 1991, 1995). However, learning a new category at the prosodic structural level requires different mechanisms than learning a new category at the segmental level—partly because similar suprasegmental features (F0, duration, and intensity) are used crosslinguistically and partly because it requires learning the prosodic system of the L2 as a whole. According to SLM, to learn a new segmental category, learners must be able to identify phones of the L2 as new before they are eventually able to establish the new category. However, when Korean learners, for example, encounter initial F0 rise, what has to be identified as new is not the F0 rise itself—because the same cue is used in Korean—but the particular new way F0 rise is aligned with segmental content in the L2 (i.e., its association with the initial syllable in the L2 as opposed to the final syllable in Korean). In this sense, learning to use prosodic cues for segmentation is more similar to learning to use specific phonetic cues for the recognition of L2 sounds. Broersma (2005, 2010), for example, showed that Dutch listeners, who have experience with the use of vowel duration as a perceptual cue for several phoneme contrasts in their L1 (including word-medial fricative voicing contrasts), had difficulty using this same cue for fricative voicing in English when the fricatives occurred in an unfamiliar position—namely word finally. Thus, not the cue itself but the use of the familiar cue for a new purpose proved difficult in L2 sound perception, which is similar to the results with prosodic cues. Because the same F0 rise is used in Korean (on the final syllable), the Korean way of using the cue is likely to interfere with its use in the L2 (on the first syllable). Cho and colleagues (2007) argued that lexical segmentation requires computation of prosodic structure, in parallel with segmental processing. They proposed, building on Salverda, Dahan, and McQueen (2003), that the prosodic structural information is computed online by the so-called prosody analyzer, using available prosodic structural cues. The results of this study suggest that learning how a particular prosodic cue is aligned with segmental content in the L2 thus requires developing a new way of parsing for L2 prosodic structure. This new L2 prosodic parser, or L2 prosody analyzer, should be distinct from the learners' L1 prosody analyzer. It is therefore important for models like SLM or PAM to take into account these structural aspects of prosody in developing the existing model or a new model, to provide a fuller description of the processing of L2 speech.

CONCLUSION

A prosodic structure that manifests how sound units are grouped together and which units are more prominent than others is an essential element of the linguistic system of any particular language. It would not be too much to say that no two languages are the same in the way the prosodic structure is built, in the way it is phonetically expressed, and in the way it is used in speech comprehension. Nevertheless, languages are extremely similar in that their prosodic structure is phonetically expressed by using a rather limited set of suprasegmental prosodic features, such as duration, F0, and intensity. These prosodic features are auditory-perceptually salient, such that they are easily and universally used in processing both native and nonnative languages. An important goal of the present study was to understand the extent to which the universally used suprasegmental prosodic features can be understood as universal or as language specific in terms of their use in nonnative lexical segmentation, and the extent to which nonnative prosodic patterns can be learned and exploited in lexical segmentation. The results of the AL learning experiment with listeners of two typologically unrelated languages, Dutch and Korean, suggest that although similar prosodic cues are used crosslinguistically due to their universally applicable auditory-perceptual saliency, the way they are used in processing an unfamiliar language is highly constrained by the way they are used in processing the listeners' L1. This is particularly true when learners are exposed to an unfamiliar language for the first time. Their dependency on the prosodic system of their L1 wanes with increased exposure to the L2, such that they learn gradually a new way of using the prosodic cues. Learning this new way, however, appears to be more complex than it might seem, as it involves learning the whole new prosodic system in which familiar prosodic cues are used in an unfamiliar way. It is our hope that these findings and discussion provide some new directions that theories of nonnative speech perception can pursue to embrace the rather complicated universal and language-specific aspects of prosody to create a more comprehensive model of L2 listening and learning.

(*Received 18 February 2011*)

NOTES

1. An anonymous *SSLA* reviewer pointed out that although prosodic structure is language specific, there is strong evidence for universal aspects of each of the prosodic dimensions, such that there is much less freedom among languages than this phrase suggests. The authors agree that the use of phonetic features is constrained by language universals, such that, for example, languages employ no more than a two-way phonological contrast using the duration feature and use a limited number of tonal patterns in forming phonological structures of intonation. Nevertheless, languages are different in the way that phonological features are phonetically realized. For example, by investigating voice onset time (VOT) patterns of 18 languages, Cho and Ladefoged (1999) observed that, even though none of those languages uses more than two categories for voiceless stops (i.e., unaspirated and aspirated), languages are different in choosing specific VOT values to mark the same phonological contrast, and that when there is only one voiceless stop category in a language, virtually any value on a VOT continuum can be chosen for it. They then argued that each language chooses an arbitrary modal VOT value for each of the phonological categories in the language. Another example can be found in the way that languages realize the same intonational category. Ladd and colleagues (2009) showed that the phonologically same pitch rise (i.e., L + H*) for stressed syllables in both English and Dutch is different in fine phonetic detail, which shows a slightly but consistently earlier pitch peak in Dutch than in English (see Arvaniti, Ladd, & Mennen, 1998). Following these studies, an important theoretical stance taken in the present study is that the fine-grained phonetic detail of phonological categories is arbitrarily determined by the language, insofar as it meets universally applicable physiological and aerodynamic constraints, and as a result it is poorly predictable by the phonology.

2. The learnability issue can also be considered in terms of specific models of L2 acquisition. In the present study, the issue is considered more generally in terms of whether a cue occurs across languages (and thus is universally available) or is specific to either the L1 or the L2. The implications for specific theories of L2 speech learning such as Flege's (1995) SLM are mentioned in the General Discussion section.

3. L = low tone, H = high tone

4. It is important to note that initial F0 rise may be formed in the first syllable either by an L + H tonal sequence in the syllable or by a H tone in the syllable preceded by a L tone in the immediately preceding syllable. In the present study, the term "F0 rise" is used insofar as the H tonal element exists in the syllable in question, but when the speech materials were manipulated to create F0 rise, a H tone was used to avoid the complex issue of how the rising slope is determined.

5. In Korean, a high pitch after the tensed or aspirated stops is not due to the simple local perturbation caused by the presence of a voiceless stop, but it is phonologized in the intonational system, such that the vowel following the tensed or aspirated stops in phrase-initial position is assigned a H tone, which is realized in the entire vowel (see Jun, 2000; Kingston & Diehl, 1994).

6. An AP in Korean may also be marked by a tonal pattern other than the phonologically defined LHLH pattern (Jun, 2000), but the frequency of those alternative patterns is negligible. Kim's (2004) study, which transcribed Korean intonation patterns in read speech and radio drama, showed that when AP-initial consonants were not aspirated or tensed, more than 90% of AP-initial multisyllabic content words started with a L tone, and about 85% of APs ended with a H tone. Listeners make use of the initial L and the final H in lexical segmentation, but not the alternative (less frequent) initial H or the final L (Kim & Cho, 2009).

7. A F0 rise pattern occurs in both Korean and Dutch not only utterance finally (as in marking yes-no questions) but also phrase finally in the middle of an utterance (as in a listing intonation).

8. Participants' L2 backgrounds may also influence the AL learning performance. However, it was difficult to control the L2 factor as well as prosodic factors in the present study, which already required 200 participants. A separate study is currently being conducted to see how the English proficiency level of Koreans influences learning an AL with and without prosodic cues used in English, in comparison with the performance of English-L1 speakers.

9. As a *SSLA* reviewer pointed out, in the absence of between-syllable coarticulation, the resulting speech materials sounded somewhat artificial and unnatural. Like in Tyler and Cutler (2009), this was, however, exactly what was intended, as unwanted coarticulation effects, which are potentially different in Korean and Dutch, would otherwise have been confounded with the prosodic effects that were tested in the present study.

10. The sound samples of each condition can be found at http://tcho.hanyang.ac.kr/AL_soundsamples.html.

11. A *SSLA* reviewer pointed out the study by Lerdahl and Jackendoff (1983), which showed that consistently alternating durational and tonal sequences such as short-short-long and long-long-short, or low-low-high (LLH) and high-low-low (HLL) are generally perceived as units in music by listeners of languages that are radically different in their rhythmic structures. Given that there is a remarkable structural resemblance between language and music, it is possible to argue that the prosodic cues that are used in the present study may be general perceptual cues, and not specific to languages. However, the results of the present study, together with those reported in Tyler and Cutler (2009), clearly demonstrate that the physically same prosodic cues are used differently by listeners of different languages. It appears, then, that, although the rhythmic cues used in music may be processed in much the same way by listeners with different language backgrounds, the cues are encoded language specifically in the linguistic system, resulting in language-specific use of the cues in processing language.

REFERENCES

Aoyama, K., Flege, J., Guion, S., Akahane-Yamada, R., & Yamada, T. (2004). Perceived phonetic dissimilarity and L2 speech learning: The case of Japanese /r/ and English /l/ and /r/. *Journal of Phonetics*, *32*, 233–250.

Arvaniti, A., Ladd, D. R., & Mennen, I. (1998). Stability of tonal alignment: The case of Greek prenuclear accents. *Journal of Phonetics*, *26*, 3–25.
Bagou, O., Fougeron, C., & Frauenfelder, U. H. (2002, April). *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.
Beckman, M. E. (1996). The parsing of prosody. *Language and Cognitive Processes*, *11*, 17–67.
Beckman, M. E., & Pierrehumbert, J. (1986). Intonational structure in Japanese and English. *Phonology Yearbook*, *3*, 255–309.
Best, C. T. (1994). The emergence of native-language phonological influences in infants: A perceptual assimilation model. In J. C. Goodman & H. C. Nusbaum (Eds.), *The development of speech perception* (pp. 167–224). Cambridge, MA: MIT Press.
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). Timonium, MD: York Press.
Best, C. T., McRoberts, G. W., & Goodell, E. (2001). Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. *Journal of the Acoustical Society of America*, *109*, 775–794.
Bohn, O. S., & Munro, M. J. (Eds.). (2007). *Language experience in second language speech learning: In honor of James Emile Flege*. Amsterdam: Benjamins.
Broersma, M. (2005). Perception of familiar contrasts in unfamiliar positions. *Journal of the Acoustical Society of America*, *117*, 3890–3901.
Broersma, M. (2010). Perception of final fricative voicing: Native and nonnative listeners' use of vowel duration. *Journal of the Acoustical Society of America*, *127*, 1636–1644.
Cambier-Langeveld, T. (2000). *Temporal marking of accent and boundaries*. Unpublished doctoral dissertation, University of Amsterdam, The Netherlands.
Cambier-Langeveld, T., Nespor, M., & van Heuven, V. J. (1997, September). *The domain of final lengthening in production and perception in Dutch*. Paper presented at Eurospeech, Rhodes, Greece.
Cho, T. (2002). *The effects of prosody on articulation in English*. London: Routledge.
Cho, T. (2004). Prosodically conditioned strengthening and vowel-to-vowel coarticulation in English. *Journal of Phonetics*, *32*, 141–176.
Cho, T., & Jun, S.-A. (2000). Domain-initial strengthening as featural enhancement: Aerodynamic evidence from Korean. *Chicago Linguistics Society*, *36*, 31–44.
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., & Ladefoged, P. (1999). Variation and universals in VOT: Evidence from 18 languages. *Journal of Phonetics*, *27*, 207–229.
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 non-native 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.
Chung, K., Chang, S., Choi, J., Nam, S., Lee, M., Chung, S., et al. (1996). *A study of Korean prosody and discourse for the development of speech synthesis/recognition system*. Daejun, Korea: KAIST Artificial Intelligence Research Center.
Cutler, A. (1994). Segmentation problems, rhythmic solutions. *Lingua*, *92*, 81–104.
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. (1995). The supraglottal articulation of prominence in English: Linguistic stress as localized hyperarticulation. *Journal of the Acoustical Society of America*, *97*, 491–504.
Dilley, L., & McAuley, J. D. (2008). Distal prosodic context affects word segmentation and lexical processing. *Journal of Memory and Language*, *59*, 294–311.
Flege, J. (1991). Perception and production: The relevance of phonetic input to L2 phonological learning. In T. Heubner & C. W. Ferguson (Eds.), *Crosscurrents in second language acquisition and linguistic theory* (pp. 249–289). Amsterdam: Benjamins.
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). Timonium, MD: York Press.
Flege, J., & MacKay, I. (2004). Perceiving vowels in a second language. *Studies in Second Language Acquisition*, *26*, 1–34.
Flege, J., & Wang, C. (1989). Native-language phonotactic constraints affect how well Chinese subjects perceive the word-final English /t/-/d/ contrast. *Journal of Phonetics*, *17*, 299–315.
Goto, H. (1971). Auditory perception by normal Japanese adults of the sounds "l" and "r." *Neuropsychologia*, *9*, 317–323.
Gussenhoven, C. (2004). Transcription of Dutch intonation. In S.-A. Jun (Ed.), *Prosodic typology: The phonology of intonation and phrasing*. Oxford: Oxford University Press.
Johnson, E. K., & Jusczyk, P. W. (2001). Word segmentation by 8-month-olds: When speech cues count more than statistics. *Journal of Memory and Language*, *44*, 548–567.
Jun, S.-A. (1993). *The phonetics and phonology of Korean prosody*. Columbus: Ohio State University Press.
Jun, S.-A. (1998). The accentual phrase in the Korean prosodic hierarchy. *Phonology*, *15*, 189–226.
Jun, S.-A. (2000). K-ToBI (Korean ToBI) labelling conventions (version 3.1). Retrieved February 18, 2011, from http://www.linguistics.ucla.edu/people/jun/ktobi/K-tobi.html.
Jun, S.-A. (Ed.). (2004). *Prosodic typology: The phonology of intonation and phrasing*. Oxford: Oxford University Press.
Kager, R. J. W. (1989). *A metrical theory of stress and destressing in English and Dutch*. Dordrecht, The Netherlands: Foris.
Keating, P. A. (2006). Phonetic encoding of prosodic structure. In J. Harrington & M. Tabain (Eds.), *Speech production: Models, phonetic processes, and techniques* (pp. 167–186). London: Psychology Press.
Keating, P. A., & Shattuck-Hufnagel, S. (2003). A prosodic view of word form encoding for speech production. *UCLA Working Papers in Phonetics*, *101*, 112–156.
Kim, S. (2004). *The role of prosodic phrasing in Korean word segmentation*. Unpublished doctoral dissertation, University of California, Los Angeles.
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., & McQueen, J. M. (2012). Phonetic richness can outweigh prosodically-driven phonological knowledge when learning words is an artificial language. *Journal of Phonetics*, *40*, 443–452.
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.
Ladd, D. R., Schepman, A., White, L., Quarmby, L. M., & Stackhouse, R. (2009). Structural and dialectal effects on pitch peak alignment in two varieties of British English. *Journal of Phonetics*, *37*, 145–161.
Lee, H. J. (1998). *Tonal realization and implementation of the accentual phrase in Seoul Korean*. Unpublished master's thesis, University of California, Los Angeles.
Lehiste, I. (1970). *Suprasegmentals*. Cambridge, MA: MIT Press.
Lerdahl, F., & Jackendoff, R. (1983). *A generative theory of tonal music*. Cambridge, MA: MIT Press.
Lively, S. E., Logan, J. S., & Pisoni, D. B. (1993). Training Japanese listeners to identify English /r/ and /l/. II: The role of phonetic environment and talker variability in learning new perceptual categories. *Journal of the Acoustical Society of America*, *94*, 1242–1255.

Logan, J. S., Lively, S. E., & Pisoni, D. B. (1991). Training Japanese listeners to identify English /r/ and /l/: A first report. *Journal of the Acoustical Society of America*, *89*, 874–886.

Mattys, S. L. (2004). Stress versus coarticulation: Toward an integrated approach to explicit speech segmentation. *Journal of Experimental Psychology: Human Perception and Performance*, *30*, 397–408.

McQueen, J. M. (1998). Segmentation of continuous speech using phonotactics. *Journal of Memory and Language*, *39*, 21–46.

McQueen, J. M. (2005). Speech perception. In K. Lamberts & R. Goldstone (Eds.), *The handbook of cognition* (pp. 255–275). London: Sage.

Pierrehumbert, J. (1980). *The phonology and phonetics of English intonation*. Unpublished doctoral dissertation, Massachusetts Institute of Technology.

Polka, L. (1991). Cross-language speech-perception in adults: Phonemic, phonetic, and acoustic contributions. *Journal of the Acoustical Society of America*, *89*, 2961–2977.

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

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.

Shattuck-Hufnagel, S., & Turk, A. E. (1996). A prosody tutorial for investigators of auditory sentence processing. *Journal of Psycholinguistic Research*, *25*, 193–247.

Shatzman, K. B., & McQueen, J. M. (2006). Prosodic knowledge affects the recognition of newly acquired words. *Psychological Science*, *17*, 372–377.

Sluijter, A. M. C., & van Heuven, V. J. (1996, October). *Acoustic correlates of linguistic stress and accent in Dutch and American English*. Paper presented at the Fourth International Conference on Spoken Language Processing, Philadelphia, PA.

Spitzer, S. M., Liss, J. M., & Mattys, S. L. (2007). Acoustic cues to lexical segmentation: A study of resynthesized speech. *Journal of the Acoustical Society of America*, *122*, 3678–3687.

Strange, W. (Ed.). (1995). *Speech perception and linguistic experience: Issues in cross-language research*. Timonium, MD: York Press.

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.

Vroomen, J., Tuomainen, J., & de Gelder, B. (1998). The roles of word stress and vowel harmony in speech segmentation. *Journal of Memory and Language*, *38*, 133–149.

Warner, N., Otake, T., & Arai, T. (2010). Intonational structure as a word boundary cue in Japanese. *Language and Speech*, *53*, 107–131.

Weber, A. (2001). *Language-specific listening: The case of phonetic sequences*. Unpublished doctoral dissertation, Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands.

Welby, P. (2007). The role of early fundamental frequency rises and elbows in French word segmentation. *Speech Communication*, *49*, 28–48.

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.