Cross-language differences in cue use for speech segmentation

Michael D. Tyler[a]
MARCS Auditory Laboratories and School of Psychology, University of Western Sydney, Locked Bag 1797, Penrith South DC, New South Wales 1797, Australia

Anne Cutler
Max Planck Institute for Psycholinguistics, Nijmegen 6500 AH, The Netherlands and MARCS Auditory Laboratories, University of Western Sydney, Locked Bag 1797, Penrith South DC, New South Wales 1797, Australia

(Received 17 September 2008; revised 24 March 2009; accepted 13 April 2009)

Two artificial-language learning experiments directly compared English, French, and Dutch listeners' use of suprasegmental cues for continuous-speech segmentation. In both experiments, listeners heard unbroken sequences of consonant-vowel syllables, composed of recurring three- and four-syllable "words." These words were demarcated by (a) no cue other than transitional probabilities induced by their recurrence, (b) a consistent left-edge cue, or (c) a consistent right-edge cue. Experiment 1 examined a vowel lengthening cue. All three listener groups benefited from this cue in right-edge position; none benefited from it in left-edge position. Experiment 2 examined a pitch-movement cue. English listeners used this cue in left-edge position, French listeners used it in right-edge position, and Dutch listeners used it in both positions. These findings are interpreted as evidence of both language-universal and language-specific effects. Final lengthening is a language-universal effect expressing a more general (non-linguistic) mechanism. Pitch movement expresses prominence which has characteristically different placements across languages: typically at right edges in French, but at left edges in English and Dutch. Finally, stress realization in English versus Dutch encourages greater attention to suprasegmental variation by Dutch than by English listeners, allowing Dutch listeners to benefit from an informative pitch-movement cue even in an uncharacteristic position. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3129127]

PACS number(s): 43.71.Hw, 43.71.Sy, 43.71.Es [AJ] Pages: 367–376

I. INTRODUCTION

Listening to continuous speech is easy in the native language and often difficult in a foreign language, and one of the reasons for this is that segmenting a continuous-speech stream into its component words encourages language-specific solutions. Among the sources of information which can help locate word boundaries are phonotactic sequencing constraints (e.g., the sequence /mg/ cannot be syllable-internal, but must contain a boundary, as in *some good*). Listeners make use of such constraints to segment speech (McQueen, 1998). In Finnish, which has word-level vowel harmony, two successive syllables containing vowels from different harmony classes must belong to different words; listeners make use of this knowledge in segmentation too (Suomi *et al.*, 1997; Vroomen *et al.*, 1998). English and Dutch are languages with variable lexical stress, but in both languages there is a strong statistical tendency for stress to fall word-initially; this too is effectively exploited by listeners in these languages (Cutler and Norris, 1988; Vroomen *et al.*, 1998).

Each of these factors is clearly language-specific. Stress placement is not a relevant factor for the many languages without stress, vowel harmony match is irrelevant in languages without vowel harmony, and though some phoneme sequence constraints (e.g., /mg/) hold across languages, many are language-specific. Sequences that cannot be syllable-internal in English (and hence must contain a boundary) may occur syllable-internally in other languages (e.g., /kv/ in German, /mr/ in Czech), and acceptable syllable-internal sequences in English may force a boundary in other languages (e.g., /ld/ as in *cold, build* would contain a boundary in German or Dutch or other languages with obligatory syllable-final obstruent devoicing). Thus the ease with which listeners segment continuous speech in their native language is in part based on efficient exploitation of the probabilities specific to that language.

Conversely, the difficulty of segmenting speech in a non-native language is in part based on unfamiliarity with that language's probabilities, and, worse, application of segmentation procedures encouraged by the native language to input for which they are inappropriate. Highly proficient German listeners to English draw on German phonotactic constraints in segmenting English (Weber and Cutler, 2006). French listeners, who can use a syllable-based segmentation procedure effectively with their native language, apply the same procedure to input in English (Cutler *et al.*, 1986) and in Japanese (Otake *et al.*, 1993), although syllabic segmentation is not used by native speakers of either of these languages. Likewise, Japanese listeners, whose native language encourages a mora-based segmentation procedure, also apply

[a] Author to whom correspondence should be addressed. Electronic-mail: m.tyler@uws.edu.au

that procedure to input in English (Cutler and Otake, 1994) and in French (Otake *et al.*, 1996), although, again, native speakers of neither language do this.

The studies showing that listeners make use of language-specific probabilities in speech segmentation have mostly used the word-spotting task (McQueen, 1998; van der Lugt, 2001; Vroomen *et al.*, 1998; Weber and Cutler, 2006). But with word-spotting, which exploits knowledge of a vocabulary, the same input cannot be presented to different language groups. This can however be achieved with artificial-language learning (ALL) techniques. In ALL studies, listeners are typically exposed for minutes on end to a continuous stream of speech made up of novel (but phonotactically acceptable) "words," and tested post-exposure on their recognition of the recurring constituent components. For instance, they might hear *pabikutibudogolatudaropitibudopabikudaropigolatu*, containing the recurring trisyllables *daropi, pabiku, golatu*, and *tibudo*; successful segmentation would enable listeners to accept these items as the words, and reject *kutibu, dogola*, and any other sequence which only occurred through juxtaposition of recurring items.

Listeners can perform this task using only the information in the transitional probability between syllables of the input (Saffran *et al.*, 1996b), and the resulting learning has been shown to generalize beyond the exposure materials (Mirman *et al.*, 2008). If the exposure materials contain useful phonetic cues, such as tiny pauses between the constituent items (Toro *et al.*, 2008), prosodic contours grouping the syllables into words (Vroomen *et al.*, 1998), or vowel harmony within words (Vroomen *et al.*, 1998, for Finnish listeners), then listeners can use these cues too. The ALL task has the further advantage that it can be used with populations without a lexicon, such as prelinguistic infants, where it has yielded valuable insights into the statistical learning capacities of young language learners (Saffran *et al.*, 1996a; Johnson and Jusczyk, 2001; Thiessen and Saffran, 2003).

This independence of lexical knowledge makes ALL studies also suitable for direct comparisons across languages. If the component items are made up of phonemes with a high cross-language frequency of occurrence, the input can be virtually language-neutral. Although ALL research has been a real growth area in recent years, and, in particular, many infant/adult comparisons have been undertaken, the primary focus has been the use of syllable-to-syllable transitional probability (TP) rather than of the language-specific information (phonotactic constraints, rhythmic structure), which, as described above, has been shown to characterize speech segmentation by adult listeners. Presumably in consequence of this, there have been remarkably few direct cross-language comparisons with ALL techniques.

This is somewhat surprising given that ALL techniques allow manipulation of segmentation cues. Vroomen *et al.* (1998) accompanied their word-spotting study on Finnish with an ALL study involving Finnish, French, and Dutch listeners, in which they manipulated two cues: vowel harmony and a cue they called stress that was realized as a fundamental frequency (f0) contour rising across the first syllable of a trisyllabic item, and then gradually decreasing to a baseline across the second and third syllables. They found differences in the use of these two cues across the three groups: Finnish listeners made use of the vowel harmony cue while the other two groups did not, and Finnish and Dutch listeners used the "stress" cue while French listeners did not. Both patterns were in agreement with the phonological facts: Finnish has vowel harmony while the other languages do not, and Finnish is a fixed stress language, Dutch is a free stress language, while French is not a stress language.

The currently available results suggest that listeners can use any segmentation cue of which they have had experience in their native language, from the universally computable cue of TP through any aspect of language-particular structure, but that they may not make use of cues with which they are unfamiliar. However, many more dimensions of this account remain to be explored. In infancy, for instance, there is evidence that TP allows discovery of the language-specific cues that will be most effective for acquiring the native vocabulary, after which the language-specific cues are used in preference to TP (Johnson and Jusczyk, 2001; Johnson and Seidl, 2009; Thiessen and Saffran, 2003). For adults, we have as yet little information about the relative strength of alternative cues; we do not know whether TP is the only type of information that is universally computable while all other cues are language-specific, or whether there are also cues which will prove to be universal; finally, we do not know whether a given cue is used in same way by all listeners who make use of it.

A prediction may easily be made concerning a candidate for a universal non-TP cue: final lengthening. It has been known for at least a century that regular sounds varying in duration tend to be heard as forming iambic sequences, that is, with the longer elements in final position (Woodrow, 1909). For linguistics, Hayes (1995) formulated the iambic/trochaic law, whereby intensity contrast produces trochaic grouping while durational contrast produces iambic grouping. Bolinger (1978) claimed that there are just two prosodic universals: the signaling of prominence and the signaling of juncture, and for the latter, Vaissière (1983) in a cross-language survey proposed that pre-boundary final lengthening is linguistically universal. Testing the iambic/trochaic law in an experiment with speech (synthetic CV syllables separated by 200 ms silent intervals) and nonspeech (square waves, similarly spaced) stimuli, Hay and Diehl (2007) found that both English- and French-speaking listeners preferred to group durationally varying sequences so as to produce an iambic rhythm. Saffran *et al.* (1996b), in the only cross-position preference ALL study with adult listeners, found that English speakers benefited from final lengthening over and above TP information, but not from initial lengthening.

A prediction may also be made concerning relative sensitivity to cues. Recent evidence from English and Dutch has revealed subtle differences in the use of the acoustic correlates of lexical stress. These two closely related languages both have variable lexical stress, and the phonological determinants of Dutch and English stress placement are virtually identical (van der Hulst, 1999). However, unstressed syllables show vowel reduction far more often in English than in Dutch; as a result of this, lexico-statistical analyses reveal that vowel quality suffices to effect distinctions between words to a greater extent in English than in Dutch, and taking suprasegmental cues to stress into account in speech recognition yields a more substantial payoff in Dutch than in English (Cutler and Pasveer, 2006). For example, *cigar* and *cigarette* have different vowels in the first syllable in most dialects of English, while the cognate words in Dutch have the same vowel; and the words *octopus* and *October* exist in both languages, but begin to differ segmentally on the fourth phoneme in English (the second vowel in English *octopus* is reduced), but only on the fifth phoneme in Dutch. Although suprasegmental cues distinguish different levels of stress in both languages, making use of these cues (in *ci-* or in *octo-*, in these examples) thus pays off more for distinguishing words in Dutch. Listeners act in accord with this, showing stronger effects of suprasegmental mismatch in word recognition in Dutch (Donselaar *et al.*, 2005) than in English (Cooper *et al.*, 2002). Indeed, in judging the source of English syllables differing only in stress level (e.g., *mus-* from *music* vs *museum*), Dutch listeners outperformed native English listeners (Cooper *et al.*, 2002), and although there were significant acoustic differences between members of these syllable pairs on all suprasegmental dimensions affected by stress, the responses of the Dutch listeners were more closely correlated with the acoustic variation than were those of the native listeners (Cutler *et al.*, 2007). English listeners' judgments of English stress are principally determined by vowel quality rather than by any suprasegmental cue (Fear *et al.*, 1995), whereas Dutch listeners' judgments of stress in the same English stimuli are more fine-grained and make better use of the suprasegmental cues (Cutler, 2009), as indeed do their stress judgments in their native language (Sluijter *et al.*, 1997). It is thus reasonable to predict that suprasegmental cues may be better exploited by Dutch than by English listeners in ALL tasks too.

In the present ALL study we use the cues most often manipulated in the one-language studies: lengthening and pitch movement. Both are suprasegmental cues familiar to all our listeners. As noted above, lengthening is associated universally with iambic structures; it is a cue to a right-edge boundary. Pitch movement is principally associated with the expression of prominence, but it exhibits no positional restrictions. We use the two cues orthogonally in right- and left-edge positions, and contrast them with no separate cue; the TP structure in all the materials is otherwise identical.

We present these stimuli to listeners from three languages: English, French, and Dutch. This comparison allows us first to examine the effects on segmentation performance of cross-language differences in preferred prosodic structure. French has more right-edge (iambic) and English and Dutch more left-edge (trochaic) boundary phenomena. In general, we therefore predict that French listeners will show greater sensitivity to cues in item-final position while English and Dutch listeners will show greater sensitivity to cues in item-initial position. Further, French has no stress while both English and Dutch have stress, and in both the latter languages stress differences have acoustic reflections in pitch movement and in duration, whereby in both, stress affects f0 more strongly than it affects duration. If the universal status of the durational cue prevails over its language-specific realizations, then a final lengthening cue would prove useful to all listeners, and to a greater extent than an initial lengthening cue (or no cue other than TP information). Initial lengthening, if it is useful as a cue, may, however, prove useful to English and Dutch listeners to a greater extent than to French listeners.

The cross-language comparison also allows us to examine the relative sensitivity of English and Dutch listeners to the cues we are manipulating, both of which are suprasegmental in nature. As described above, Dutch listeners have been shown to display greater sensitivity to suprasegmental cues to stress in their own language than English listeners do in theirs and also to be more sensitive than English listeners to the suprasegmental cues to stress which English offers. We predict that if differences appear in how the cues are used by these two listener groups with prosodically highly similar native phonologies, then the differences will be in the direction of greater exploitation of the cues we provide by the Dutch listeners than by the English.

II. EXPERIMENT 1: VOWEL LENGTHENING CUES

A. Method

1. Participants

In each of the three language groups, 24 participants were randomly assigned to each cue condition (TP-only, left-edge cue, and right-edge cue; $n=24\times 9=216$). French participants were psychology students at the Université de Bourgogne, Dijon, France. All had acquired French from birth, with the exception of one participant in the TP-only condition who acquired French at the age of 3 (Arabic first language [L1]), and all but eight participants had learned some English at school. Around half of the participants had also learned Spanish. The three conditions (TP-only, left-edge cue, right-edge cue) were matched as closely as possible: each condition contained 21 female participants, and the mean ages were, respectively, 19.25 (s.e.m. 0.34), 18.92 (s.e.m. 0.21), and 19.33 (s.e.m. 0.34) years. Dutch participants were recruited from the participant panel at the Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands, and were university students from a variety of academic disciplines. All were native speakers of Dutch, all had learned English at school; the majority had also learned German, and two-thirds had learned French. In the three conditions, there were, respectively, 18, 18, and 19 females, and the mean ages were 21.83 (s.e.m. 0.60), 22.04 (s.e.m. 0.56), and 21.00 (s.e.m. 0.45) years. The English speakers were first-year psychology students at the University of Western Sydney, Australia. All were native speakers of English, and there was no systematic pattern of exposure to other languages in the sample. In the three conditions there were, respectively, 19, 21, and 22 females, and the mean ages were 23.17 (s.e.m. 1.13), 24.42 (s.e.m. 1.77), and 21.13 (s.e.m. 0.88) years.

2. Stimulus materials

The artificial language consisted of the concatenation of nine words. In ALL studies the words usually have an equal number of syllables (e.g., three: Saffran *et al.*, 1996b; Vroomen *et al.*, 1998). However, if that were the case here, the isochronous rhythm created by the addition of a vowel lengthening cue could improve participants' performance over and above any influence of the cue itself. The words of the language used here therefore varied in length—six trisyllabic words and three words of four syllables. The 30 consonant-vowel syllables were constructed by exhaustively combining six consonants and five vowels that occur in the phoneme inventories of French, Dutch, and English: /p, b, m, f, s, k/, and /a, i, ɛ, ɔ, u/.

There is always the possibility that listeners may come to an ALL task with biases that influence word-boundary detection during the exposure phase (Reber and Perruchet, 2003). For example, certain syllables, consonants, or vowels may occur with a higher probability at word boundaries in the listener's native language, and combinations of syllables in the artificial language may resemble real words. To counteract such effects, we randomly allocated the 30 syllables to words, without repetition, such that the nine words of the artificial language were composed of a different unique combination of syllables for each of the 24 participants in any given condition. For each combination, TP-only, left-edge-cue, and right-edge-cue versions were created; thus each combination was presented to one participant per condition, and the conditions were balanced for variation across syllable combinations.

The 24 × 3 languages were generated using the MBROLA diphone synthesizer (Dutoit *et al.*, 1996). Each consonant and vowel was assigned a base length of 116 ms, resulting in a syllable length of 232 ms (following Peña *et al.*, 2002). As a partial control for effects of phonetic differences between realizations of the vowels and consonants across the three languages, and effects of native-language experience, half of the participants in each language group heard a language synthesized with a male French voice (MBROLA's fr1 diphone database) and the other half a male Dutch voice (the n12 diphone database). There is currently no Australian English diphone database for MBROLA. Note that each phoneme was coarticulated with the following phoneme, regardless of its position in the word.

The f0 was set to a monotone 120 Hz in all three cue conditions, and in the cued conditions the vowel in either the first syllable (left-edge cue) or last syllable (right-edge cue) of each statistical word was lengthened by 60 ms. The initial exposure lasted for a total of 11–12 min., depending on the condition, and was divided into five blocks of equal length to help maintain participants' attention. A 5-s fade-in and fade-out was applied to each block so that participants would not have access to word-boundary cues from the beginning and end of the sequence. The words were presented 19 times per block and in random order, with the sole constraint that a given word could not follow itself.

The test phase included 27 pairs of items. One member of each pair was a word from the language and the other was a part-word, that is, a sequence of syllables that occurred in the stream but crossed a word boundary. For half of the participants, the part-words were formed from the last two (or three) syllables of a word and the first syllable of another, and for the other half they were formed from the last syllable of a word and the first two (or three) syllables of another. All nine words of the language were used in the test phase, along with nine part-words. Each word was paired with three part-words and each part-word was paired with three words to counteract learning during the test phase. All items in the test phase were presented with the vowel lengthening cue corresponding to the stimulus condition (either TP only, left- or right-edge), following Vroomen *et al.* (1998).[1] The words and part-words were separated by an interstimulus interval of 500 ms. The order of words and part-words in the item pairs was counterbalanced, and the item pairs were presented in random order. The five exposure blocks and 27 pairs of test items were presented over headphones using a computer.

3. Procedure

Participants were tested in groups of one, two, or three, each seated in front of a different computer. Verbal instructions were given by the first author, who interacted with the French participants in French and with the Dutch and Australian participants in English. To ensure that all of the directions were understood, written instructions were also provided in the participant's native language.

Participants were instructed that they would hear an artificial language, consisting of a sequence of syllables with no pauses; they were asked to pay attention to the exposure stream without reflecting too much on what they were hearing, or trying to guess the purpose of the experiment. If they felt their attention start to wander they were to try to focus again on the task. Participants were made aware that there would be a test phase after the exposure phase.

Before the test phase, the participants were told that the artificial language consisted of nonsense strings that the experimenter had designated to be the words of the language. The purpose of the test phase was to find out if they had learned anything about those words during exposure. It was stressed that they would not be real words in the participant's native language, and that any resemblance to known words would be coincidental. Participants listened to each pair of a word and a part-word and indicated, by pressing the keys "1" or "2," whether the word of the language was the first or second member of the pair. They were told to guess if unsure.

B. Results and discussion

Percent correct scores were calculated for each participant, and then each group's mean score was derived from these values. The mean percent correct responses for each language group in each of the three cue conditions are presented in Fig. 1, and the exact values and standard errors of the mean are presented in Table I.[2] An alpha level of 0.05 was used for all statistical tests unless otherwise specified.

Before assessing whether word segmentation improved in the edge cue conditions, relative to TP-only, it is necessary first to test whether participants segmented the artificial language in the TP-only condition. If participants did not segment the artificial speech stream, then their performance on the two-alternative forced-choice test would not be greater

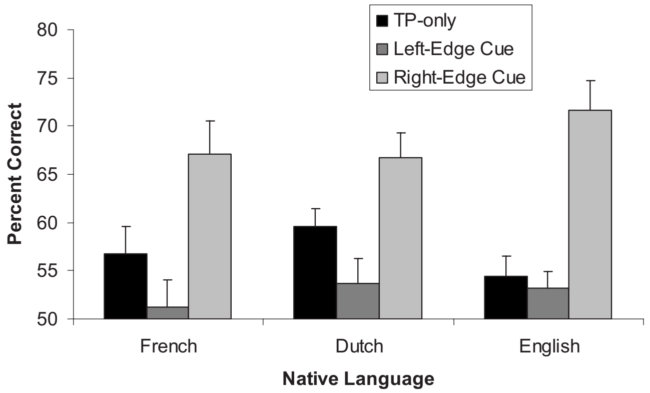

FIG. 1. Mean percent correct scores in the test phase for each language group in each cue condition in Experiment 1 (vowel-lengthening cues). Error bars represent s.e.m.

than chance (50%). For reference, the left-edge and right-edge cue conditions were also analyzed. Results of one-sample *t*-tests against a chance score of 50% are shown in Table I. Participants from each language group performed above chance in the TP-only and right-edge-cue conditions, but none of the language groups performed above chance in the left-edge-cue condition.

Having established that the artificial language can be segmented on the basis of TP cues only, our second analysis used planned contrasts to test whether segmentation was affected by cue (i.e., left-edge or right-edge vowel lengthening) and whether that varied according to the listener's native-language background. The *left-edge difference* and the *right-edge difference* contrasts compared the baseline TP-only condition with the left-edge and right-edge cue conditions, respectively. As these contrasts are not orthogonal, a Bonferroni correction was applied to the analysis (see Betz and Levin, 1982). Two additional planned contrasts assessed the influence of the listener's native language on learning. To test for effects of native-language prosodic preferences, the *language type* contrast compared the scores of French listeners (iambic, no-stress) with the combined scores of Dutch and English listeners (trochaic, with stress). The *stress language prosodic sensitivity* contrast compared scores of Dutch and English listeners only, to test performance of listeners who are more sensitive to suprasegmental information (i.e., Dutch) versus those who are less sensitive (i.e., English). Four interaction contrasts were also calculated to test for differential effects of cue location as a function of language background.

The results of the planned contrast analysis are shown in Table II. The only significant contrast was right-edge difference, with participants' scores being significantly higher in the right-edge than TP-only condition. Although performance in the left-edge condition dropped to chance level, the score was not significantly lower than the TP-only condition. None of the interaction contrasts was significant.

Experiment 1 showed, therefore, that all participants, regardless of language background, benefited from vowel lengthening if and only if it was a right-edge cue. Cross-language differences in preferred prosodic structure had no effect. This is consistent with a universal status for final lengthening as a boundary cue.

III. EXPERIMENT 2: PITCH-MOVEMENT CUES

A. Method

1. Participants

Another 216 participants, 72 from each of the same populations as for Experiment 1, took part in Experiment 2; again 24 were randomly assigned to each cue condition. All French participants had acquired French from birth, with the exception of one participant in the TP-only condition (L1: Mauritian creole), and two in the right-edge-cue condition (L1: Mandinka, Portuguese) who all acquired French in early childhood. All had also learned some English at school, around half had learned Spanish, and around one-third had learned German. In the TP-only, left-edge-cue and right-edge-cue conditions, respectively, there were 20, 18, and 18 females, and the mean ages were 20.82 (s.e.m. 0.76), 21.17 (s.e.m. 1.71), and 20.33 (s.e.m. 0.50) years. Dutch participants were native speakers of Dutch, had learned English at school, and had also been exposed to some French at school. In the three conditions, there were, respectively, 19, 19, and 20 females, and the mean ages were 20.63 (s.e.m. 0.49), 20.13 (s.e.m. 0.39), and 20.59 (s.e.m. 0.65) years. The English participants were again all native speakers of that language; 12 had taken some French at school. In the three conditions there were, respectively, 16, 18, and 20 females, and the mean ages were 20.29 (s.e.m. 1.11), 20.04 (s.e.m. 0.95), and 20.96 (s.e.m. 1.26) years.

2. Stimuli and procedure

The artificial languages used in this experiment had the same structure, and were constructed in the same manner as those used in Experiment 1, with the exception of the fundamental frequency characteristics and the vowel length. Syllable duration was that of the TP-only condition in Experiment 1 (232 ms). The f0 was set to a monotone 120 Hz for all syllables in the TP-only condition, while in the cued con-

TABLE I. Mean percent correct scores, standard error of the mean, and *t*(23) values from a one-sample *t*-test against chance (50%) for Experiment 1 (vowel-lengthening cues). Values marked with an asterisk were significant at the 0.05 level.

	TP-only			Left-edge cue			Right-edge cue		
Native Language	*M*	s.e.m	*t*(23)	*M*	s.e.m	*t*(23)	*M*	s.e.m	*t*(23)
French	56.79	2.76	2.46*	51.23	2.78	0.44	67.13	3.39	5.05*
Dutch	59.57	1.84	5.21*	53.70	2.60	1.42	66.67	2.63	6.34*
English	54.48	2.06	2.17*	53.24	1.64	1.98	71.60	3.11	6.95*

TABLE II. Planned contrast analysis for Experiment 1 (vowel-lengthening cues). Contrast values are percent mean difference scores, and asterisks indicate significance with Bonferroni adjustment.

				Bonferroni 95% confidence interval	
Contrast	*F*(1,207)	Contrast value	s.e.m.	Lower	Upper
Language type	0.66	1.49	1.83	−2.65	5.63
Prosodic sensitivity	0.01	−0.21	2.12	−4.99	4.57
Left-edge difference	3.97	−4.22	2.12	−9.00	0.56
Right-edge difference	29.62*	11.52	2.12	6.74	16.30
Language type × left-edge difference	0.20	−2.01	4.49	−13.32	9.31
Language type × right-edge difference	0.16	−1.78	4.49	−13.09	9.54
Prosodic sensitivity × left-edge difference	0.80	−4.63	5.19	−17.70	8.44
Prosodic sensitivity × right-edge difference	3.74	−10.03	5.19	−23.10	3.04

ditions a parabolic f0 contour with its peak at 170 Hz was imposed on the cued syllable (following Thiessen and Saffran, 2003). Procedure was as in Experiment 1.

B. Results and discussion

Mean percent correct responses for each language group in each of the three conditions of Experiment 2 are displayed in Fig. 2, and the exact values and standard errors of the mean are listed in Table III.[3] As can be seen, all participants performed above chance in the TP-only condition. The pattern of learning across conditions was similar to Experiment 1 for French listeners, but a different pattern emerged here for Dutch and English listeners. Dutch listeners performed above chance in all conditions, whereas English listeners performed above chance in all but the right-edge-cue condition.

Planned contrasts were applied to the data as in Experiment 1, and the results are shown in Table IV. As in Experiment 1, the significant right-edge difference contrast shows that, across the three listener groups, participants scored higher in the right-edge condition than the TP-only condition. (That is, the magnitude of the difference for French and Dutch listeners combined was sufficient to compensate for the English listeners' chance performance in the right-edge condition.) The significant Language type × left-edge difference interaction contrast shows that stress-language listeners (Dutch and English) benefited from the left-edge cue more than the French listeners. The prosodic sensitivity × right-edge difference contrast qualifies the significant overall right-edge difference by showing that Dutch listeners benefited more from the right-edge cue than English listeners. These analyses confirm the pattern of results seen in Fig. 2—French listeners benefit from right-edge cues only, Dutch listeners benefit from both left- and right-edge cues, whereas English listeners benefit from left-edge cues only.

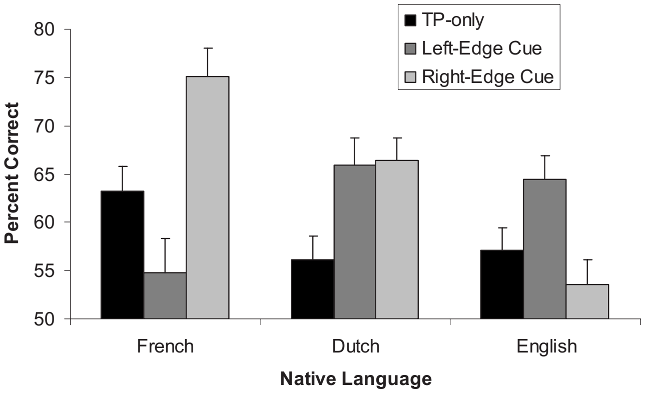

FIG. 2. Mean percent correct scores in the test phase for each language group in each cue condition in Experiment 2 (pitch-movement cues). Error bars represent s.e.m.

Experiment 2 has thus shown that pitch-movement cues to segmentation produce different results with listeners from different languages. In contrast to the universally consistent pattern revealed for vowel lengthening cues in Experiment 1, sensitivity to pitch-movement cues to segmentation is dependent on language-specific factors: preferred prosodic structure and patterns of stress realization.

IV. CONCLUSIONS

ALL as a tool for investigating speech segmentation reveals both universal similarities and cross-linguistic differences. Consistent with our predictions, based on the linguistic literature and one prior result with English-speaking listeners, lengthening realized in final position proved a powerful cue, significantly improving the segmentation performance of listeners from all language backgrounds we tested. Also consistent with our predictions, based on the cross-language differences in prosodic structure, listeners whose languages exhibit a preference for trochaic stress benefited more from a left-edge pitch cue, while listeners whose language displays preferred iambic prominence benefited more from the same cue realized in right-edge position. Finally, our prediction concerning the two highly similar stress languages, Dutch and English, was also supported: Dutch listeners exploited the presence of a pitch cue to a greater extent than English listeners did.

Thus exactly the same artificial-language input can be parsed differently if experience with the native language encourages attention to different cues in the input. Even when languages encourage use of the same cue in the same position, this does not entail that the way in which the cue is used is also the same; it too can differ. However, there are also cues that appear to be used in the same way across prosodically different languages.

Vowel lengthening is a powerful segmentation cue if it is associated with the right edge of structural components. The

TABLE III. Mean percent correct scores, standard error of the mean, and $t(23)$ values from a one-sample t-test against chance (50%) for Experiment 2 (pitch-movement cues). Value marked with an asterisk were significant at the 0.05 level.

	TP-only			Left-edge cue			Right-edge cue		
Native language	*M*	s.e.m	*t*(23)	*M*	s.e.m	*t*(23)	*M*	s.e.m	*t*(23)
French	63.27	2.50	5.31*	54.78	3.55	1.35	75.15	2.93	8.59*
Dutch	56.17	2.43	2.54*	65.90	2.90	5.48*	66.36	2.39	6.84*
English	57.10	2.33	3.05*	64.51	2.36	6.15*	53.55	2.52	1.41

widespread use of final-element lengthening suggests that it does not derive solely from linguistic structure but is more general in application (consider that lengthening of units in final position is also observed in music: Lindblom, 1978; Palmer, 1997). Hay and Diehl (2007) interpreted their finding of similar grouping preferences in speakers of French and English as supporting an explanation of the iambic/trochaic law in terms of general auditory mechanisms rather than linguistically based regularities. Our present findings are fully consistent with a general explanation of this nature.

Both the studies of Hay and Diehl (2007) and of Thiessen and Saffran (2003) were apparently conceived in the expectation that English listeners might prefer, or be more sensitive to, left-edge lengthening rather than right-edge lengthening. In both cases the authors cited the results showing that English listeners use stress in segmentation, combined with the fact that lengthening is a correlate of English stress. However, stress in English is multiply determined (see Cutler, 2005, for a review); there are suprasegmental cues in f0, duration, and amplitude, but stress placement is most highly correlated with segmental structure: syllable weight and vowel quality. Furthermore, the literature clearly shows that for lexical-level segmentation, English listeners mainly rely on the segmental information. Both in natural listening and in the laboratory, their preferred segmentation strategy is to postulate a word boundary at the onset of any strong syllable, that is, any syllable containing a full vowel (Cutler and Butterfield, 1992; Cutler and Norris, 1988). This heuristic pays off very effectively in segmenting typical English speech (Cutler and Carter, 1987), and has correctly accounted for the findings on segmentation of English when it is incorporated into computational models of word recognition in continuous speech (Norris *et al.*, 1997; Norris and McQueen, 2008). The literature also shows that English listeners make less use of suprasegmental cues for lexical processing than the acoustic structure of speech supports (Fear *et al.*, 1995; Cooper *et al.*, 2002), although the use of durational cues for syntactic processing is robust (Scott, 1982). Thus the previous findings and our current findings are fully consistent.

Lexical segmentation in normal speech recognition operates as efficiently and as rapidly as listeners can manage. Current models of spoken-word recognition (e.g., Norris and McQueen, 2008) assume that listeners evaluate multiple lexical hypotheses concurrently, and, in fact, this process itself parses continuous input into a sequence of words; nevertheless, listeners exploit the further cues to segmentation which speech signals provide, at many different levels, as ample empirical evidence attests (Mattys *et al.*, 2005). Explicit segmentation using cues in the signal, and segmentation arising from concurrent evaluation of word hypotheses, can moreover be shown to be distinct processes. Cutler and Butterfield (1992) analyzed a corpus of errors of segmentation in English, in which the predicted effect of vowel quality was found: a syllable was more likely to be erroneously interpreted as word-initial if it contained a full vowel. Effects of lexical structure on segmentation, however, were not observed; thus syllables were not more often taken as word-initial when they began more words, and erroneously reported words were more frequent than the actual words in the input only when boundaries were correct, not when boundaries had been inferred from the vowel quality cue. This pattern is consistent with use of cues in the input to deliver an initial segmentation which then constrains the set of lexical hypotheses for evaluation.

Silence abutting a word would putatively be the strongest of all such cues, and even tiny silences inserted between words produce significant improvement in ALL performance

TABLE IV. Planned contrast analysis for Experiment 2 (pitch-movement cues). Contrast values are percent mean difference scores, and asterisks indicate significance with Bonferroni adjustment.

				Bonferroni 95% confidence interval	
Contrast	*F*(1,207)	Contrast value	s.e.m.	Lower	Upper
Language type	4.03	−3.81	1.90	−8.09	0.48
Prosodic sensitivity	4.08	−4.42	2.19	−9.37	0.52
Left-edge difference	1.73	2.88	2.19	−2.07	7.83
Right-edge difference	7.94*	6.17	2.19	1.23	11.12
Language type × left-edge difference	13.46*	−17.05	4.65	−28.76	−5.34
Language type × right-edge difference	3.40	8.57	4.65	−3.15	20.28
Prosodic sensitivity × left-edge difference	0.19	2.32	5.37	−11.21	15.84
Prosodic sensitivity × right-edge difference	6.55*	13.73	5.37	0.21	27.26

(Toro *et al.*, 2008). Although vowel lengthening as a right-edge cue most strongly signals syntactic boundaries (Klatt, 1975), even in doing so it would still function as a cue to the end of a word as well as of the phrase of which the word was a part. However, in fact, there is also consistent lengthening associated with the right edge of words (Beckman and Edwards, 1990). All this makes vowel lengthening a right-edge cue of considerable power, and listeners make appropriate use of it, as the results of Experiment 1 showed. It is even noteworthy that the listeners in all groups in Experiment 1 performed slightly (albeit not significantly) worse than with TP alone when the vowel lengthening cue was associated with the left edge of words. This may also be due to the power of the cue as a signal of right edges, but here placed, under such an interpretation, in conflict with the TP cues.

The appearance of a consistent pattern across languages in one experiment, however, does not mean that cross-language differences in prosodic structure exert no influence on ALL segmentation. The relative strength of the cues in different positions clearly varied across the three listener groups we tested, most clearly in Experiment 2 in which we manipulated pitch movement. In that experiment, French listeners gained no benefit at all from a left-edge cue and English listeners gained no benefit at all from a right-edge cue. This is exactly in accord with the preferred expression of syllabic prominence (left edge of lexical units in English, right edge of lexical units—strictly speaking, of clitic groups—in French).

Note that the pitch-movement cue as we (and others) have manipulated it is something of a caricature of what pitch does in stress variation. However, undeniably it is the case that whatever it does, it does it in characteristically different locations in English and in French, and precisely that predicted difference turned up in our results.

The pitch cue that we used was realized locally on a left- or right-edge syllable. Note that this has not always been the case when pitch cues have been manipulated in ALL studies of segmentation. For example, the pitch cue manipulated by Vroomen *et al.* (1998) (referred to by those authors as “stress”) was actually a word-level prosodic contour grouping all three syllables of the words they used into a consistent prosodic shape, with prominence on the initial syllable. It seems that this should have been a relatively easy cue to use in an ALL experiment, so that their finding that French listeners did not benefit from such a cue is somewhat surprising. Vroomen *et al.* (1998) interpreted the finding in terms of language-specific prosodic structure; the characteristic prosodic shape of French words does not correspond to the prosodic shape they used as a cue. However, in a replication of their experiment, Tyler (2006) demonstrated that French listeners indeed showed significant benefit from such a prosodic shape cue, both in learning a more complex artificial language (analogous to the ones used in the present study) and in learning from the very materials used in the Vroomen *et al.*, 1998 study.[4]

Interestingly, in one ALL study in which a pitch cue was manipulated at the syllabic level, as we did, no benefit accrued for listeners. This was a study with Spanish listeners conducted by Toro-Soto *et al.* (2007). The Spanish lexicon has a strong preponderance of words with penultimate stress, and Toro-Soto *et al.* (2007) tested the value of such a stress cue (realized as an increase in syllable pitch on the penultimate syllable of trisyllabic words). Listeners performed no better than chance with this cue (and significantly worse than their performance with TP information only, or with an initial or final stress cue of the same kind). Thus for a cue to be useful for segmenting an artificial language, it seems that it should preferably be aligned with word edges.

Finally, we also observed in Experiment 2 a significant difference, again as we had predicted, in the learning performance of our English and our Dutch participants. The Dutch listeners profited from the pitch-movement cue in both initial position and final position, while the English listeners showed a benefit only in initial position. In final position, the Dutch listeners’ performance was significantly better than that of the English listeners; the latter actually performed somewhat worse in this condition than with TP-only, again consistent with an effect which was powerfully unidirectional and in conflict, under the preferred interpretation, with TP.

The Dutch listeners, however, were able to override such a preference and make use of pitch movement when it uncharacteristically provided a consistent right-edge cue. Thus in this experiment, as in preceding studies (Cooper *et al.*, 2002; Donselaar *et al.*, 2005; Cutler *et al.*, 2007; Cutler, 2009), Dutch listeners displayed greater sensitivity than English listeners to the suprasegmental cues to stress in speech, especially, in this case, the pitch-movement cue; they could learn to use it as a marker of lexical identity irrespective of its position in words, while the English listeners used it only in the position in which it most commonly occurs in English.

We interpret this not as a reflection of cross-language differences in positional marking of stress; this is overwhelmingly initial in both English [for which Cutler and Carter (1987), reported 90% of the lexical vocabulary to be strong-initial] and Dutch (for which Schreuder and Baayen (1994), reported 87.5%). Rather, it provides yet further evidence for the greater sensitivity of Dutch listeners to suprasegmental structure. We note here that a finding by Thiessen and Saffran (2004) in an ALL experiment can also be interpreted in this light. Thiessen and Saffran (2004) observed that infants could base speech segmentation on the exploitation of spectral tilt (the relative distribution of amplitude across the spectrum, a strong cue to stress in that stressed syllables show significantly greater amplitude in the higher spectral regions than unstressed syllables). Adult English-speaking listeners, however, ignored variation in spectral tilt unaccompanied by other stress cues. Spectral tilt in isolation has been shown to be an effective cue to stress for Dutch listeners (Sluijter *et al.*, 1997) but not for English listeners (Campbell and Beckman, 1997; see Cutler, 2005, for further discussion).

This consistent pattern of findings underlines how powerful are the distributional statistics of language-specific phonology in determining listeners’ attention to speech cues, and how efficiently adult listening exploits native-language probabilities. The widespread occurrence of vowel reduction in unstressed syllables in English has encouraged English listeners to skip attention to suprasegmental variation for lexical identification, since vowel quality will virtually always provide all the information that the suprasegmental variation encodes, and vowel quality must be attended to anyway. The frequent occurrence in Dutch of unstressed syllables with full vowels (such as the first syllables of *sigaar* "cigar" or *Oktober* "October") has made Dutch listeners realize that they can profit from attending to suprasegmental variation, because they can distinguish more rapidly between potential words than would be possible on the basis of segmental information alone.

ALL techniques have enabled us to observe the effect of these different language-specific patterns, and at the same time to appreciate the strength of language-universal effects, because they have allowed a direct comparison of the use of speech cues in exactly the same input across listeners with different native languages. We have seen that listeners are very good at segmentation on the basis of distributional information alone (the TP-only conditions, in which listeners could only identify recurring items on the basis of their sequential probabilities, consistently produced above-chance performance from all listener groups). Beyond that, they can also make use of other cues where these are available. But not all cues are equal. A durational cue is a significant help, but only when it is in the universally preferred right-edge position. A pitch-movement cue is also a significant help, but for most listeners such a cue is only helpful in the characteristic native-language position—right-edge for French listeners and left-edge for English listeners. Only Dutch listeners, whose language encourages careful attention to suprasegmental information, displayed sufficient flexibility to exploit a consistent pitch-movement cue both in its expected position and in an uncharacteristic mapping.

Speech segmentation is one of the most useful language processing skills. It develops early in life, and it directly assists language learning: facility with speech segmentation in the first year of life is associated with enhanced vocabulary development in the following years (Newman *et al.*, 2006). From the earliest stages it is adapted to the native language and exploits the distributional probabilities of the input (McQueen, 1998; van der Lugt, 2001). This of course has the inevitable downside that listening to a non-native language is rendered more difficult where the structure of the native language encourages segmentation procedures inappropriate for the other language (Weber and Cutler, 2006). However, this is apparently a small price to pay for the streamlined efficiency with which native input can be divided into its component words. Even with the impoverished nonword input on offer in an ALL experiment, this efficiency can be seen in action. In ALL, listeners know the input is not real language, and they know that their only task is to use the information in the input to the best of their ability. Nonetheless, they succeed in using cues only to the extent that the cue position, and/or the cue type, coincides with their native-language experience. The artificial nature of ALL input allows these effects to be observed, in that it provides a direct window onto cross-language differences.

ACKNOWLEDGMENTS

The first author was supported by a postdoctoral fellowship from the Conseil Régional de Bourgogne, France, the NWO SPINOZA project "Native and Non-native Listening" (A.C.), and NIH Grant No. DC00403 (PI: C. Best). The authors thank Pierre Perruchet for his time and advice, especially with design and generation of the ALL stimuli; Kylie Tyler, Marloes van der Goot, Eelke Spaak, Jess Hartcher-O'Brien, and Susan Wijngaarden for research assistance; and Nigel Nettheim, Kate Stevens, and the reviewers for helpful comments on the manuscript.

[1] Note that Bagou *et al.* (2002) found no difference between cued and uncued test items in a similar task.

[2] There was no effect of synthesis voice (French or Dutch), and this factor did not interact with any other variable. All analyses reported here therefore collapse across that counterbalancing factor.

[3] As in Experiment 1, there was no effect of synthesis voice (French or Dutch), and this factor did not interact with any other variable, so analyses are again collapsed across that counterbalancing factor.

[4] The difference in results was attributed to audio presentation differences (headphones in Tyler 2006, and loudspeakers in Vroomen *et al.* 1998).

Bagou, O., Fougeron, C., and Frauenfelder, U. H. (**2002**). "Contribution of prosody to the segmentation and storage of 'words' in the acquisition of a new minilanguage," in *Proceedings of Speech Prosody 2002*, edited by B. Bel, and I. Marlien (Association pour la promotion de la phonétique et de la linguistique, Aix-en-Provence, France), pp. 159–162.

Beckman, M. E., and Edwards, J. (**1990**). "Lengthenings and shortenings and the nature of prosodic constituency," in *Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech*, edited by J. Kingston, and M. E. Beckman (Cambridge University Press, Cambridge), pp. 152–178.

Betz, M. A., and Levin, J. R. (**1982**). "Coherent analysis-of-variance hypothesis testing strategies: A general approach," J. Educ. Stat. **7**, 193–206.

Bolinger, D. L. (**1978**). "Intonation across languages," in *Universals of Human Language*, Phonology, Vol. **2**, edited by J. H. Greenberg (Stanford University Press, Stanford), pp. 471–524.

Campbell, N., and Beckman, M. E. (**1997**). "Stress, prominence, and spectral tilt," in *Intonation: Theory, Models, and Applications (Proceedings of a European Speech Communication Assoc. Workshop, September 18–20, 1997)*, edited by A. Botinis, G. Kouroupetroglou, and G. Carayiannis (ESCA and University of Athens Department of Informatics, Athens), pp. 67–70.

Cooper, N., Cutler, A., and Wales, R. (**2002**). "Constraints of lexical stress on lexical access in English: Evidence from native and non-native listeners," Lang Speech **45**, 207–228.

Cutler, A. (**2005**). "Lexical stress," in *The Handbook of Speech Perception*, edited by D. B. Pisoni, and R. E. Remez (Blackwell, Oxford), pp. 264–289.

Cutler, A. (**2009**). "Greater sensitivity to prosodic goodness in non-native than in native listeners (L)," J. Acoust. Soc. Am. **125**, 3522–3525.

Cutler, A., and Butterfield, S. (**1992**). "Rhythmic cues to speech segmentation: Evidence from juncture misperception," J. Mem. Lang. **31**, 218–236.

Cutler, A., and Carter, D. M. (**1987**). "The predominance of strong initial syllables in the English vocabulary," Comput. Speech Lang. **2**, 133–142.

Cutler, A., and Norris, D. (**1988**). "The role of strong syllables in segmentation for lexical access," J. Exp. Psychol. Hum. Percept. Perform. **14**, 113–121.

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

Cutler, A., and Pasveer, D. (**2006**). "Explaining cross-linguistic differences in effects of lexical stress on spoken-word recognition," in *Proceedings of the Third International Conference on Speech Prosody*, edited by R. Hoffman, and H. Mixdorff (TUD, Dresden), pp. 250–254.

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

Cutler, A., Wales, R., Cooper, N., and Janssen, J. (**2007**). "Dutch listeners' use of suprasegmental cues to English stress," in *Proceedings of 16th*

International Congress of Phonetic Sciences, edited by J. Trouvain and W. J. Barry (Saarbrücken, Germany), pp. 1913–1916.

van Donselaar, W., Koster, M., and Cutler, A. (**2005**). "Exploring the role of lexical stress in lexical recognition," Q. J. Exp. Psychol. A **58A**, 251–273.

Dutoit, T., Pagel, V., Pierret, N., Bataille, F., and van der Vrecken, O. (**1996**). "The MBROLA project: Towards a set of high quality speech synthesizers free of use for non commercial purposes," in *Proceedings of the Fourth International Conference on Spoken Lang. Processing*, edited by H. T. Bunnell, and W. Idsardi, pp. 1393–1396.

Fear, B. D., Cutler, A., and Butterfield, S. (**1995**). "The strong/weak syllable distinction in English," J. Acoust. Soc. Am. **97**, 1893–1904.

Hay, J. S. F., and Diehl, R. L. (**2007**). "Perception of rhythmic grouping: Testing the iambic/trochaic law," Percept. Psychophys. **69**, 113–122.

Hayes, B. (**1995**). *Metrical Stress Theory: Principles and Case Studies* (University of Chicago Press, Chicago).

Johnson, E. K., and Jusczyk, P. W. (**2001**). "Word segmentation by 8-month-olds: When speech cues count more than statistics," J. Mem. Lang. **44**, 548–567.

Johnson, E. K., and Seidl, A. H. (**2009**). "At 11 months, prosody still outranks statistics," Dev. Sci. **12**, 131–141.

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

Lindblom, B. (**1978**). "Final lengthening in speech and music," in *Nordic Prosody*, edited by E. Gårding, G. Bruce, and R. Bannert (Department of Linguistics, Lund University, Lund, Sweden), pp. 85–100.

Mattys, S. L., White, L., and Melhorn, J. F. (**2005**). "Integration of multiple speech segmentation cues: A hierarchical framework," J. Exp. Psychol. Gen. **134**, 477–500.

McQueen, J. M. (**1998**). "Segmentation of continuous speech using phonotactics," J. Mem. Lang. **39**, 21–46.

Mirman, D., Magnuson, J. S., Graf Estes, K., and Dixon, J. A. (**2008**). "The link between statistical segmentation and word learning in adults," Cognition **108**, 271–280.

Newman, R., Bernstein Ratner, N., Jusczyk, A. M., Jusczyk, P. W., and Dow, K. A. (**2006**). "Infants' early ability to segment the conversational speech signal predicts later language development: A retrospective analysis," Dev. Psychol. **42**, 643–655.

Norris, D., and McQueen, J. M. (**2008**). "Shortlist B: A Bayesian model of continuous speech recognition," Psychol. Rev. **115**, 357–395.

Norris, D., McQueen, J. M., Cutler, A., and Butterfield, S. (**1997**). "The possible-word constraint in the segmentation of continuous speech," Cognit Psychol. **34**, 191–243.

Otake, T., Hatano, G., Cutler, A., and Mehler, J. (**1993**). "Mora or syllable? Speech segmentation in Japanese," J. Mem. Lang. **32**, 258–278.

Otake, T., Hatano, G., and Yoneyama, K. (**1996**). "Speech segmentation by Japanese listeners," in *Phonological Structure and Language Processing: Cross-Linguistic Studies*, edited by T. Otake, and A. Cutler (Mouton de Gruyter, Berlin), pp. 183–201.

Palmer, C. (**1997**). "Music performance," Annu. Rev. Psychol. **48**, 115–138.

Peña, M., Bonatti, L. L., Nespor, M., and Mehler, J. (**2002**). "Signal-driven computations in speech processing," Science **298**, 604–607.

Reber, R., and Perruchet, P. (**2003**). "The use of control groups in artificial grammar learning," Q. J. Exp. Psychol. A **56A**, 97–115.

Saffran, J. R., Aslin, R. N., and Newport, E. L. (**1996a**). "Statistical learning by 8-month-old infants," Science **274**, 1926–1928.

Saffran, J. R., Newport, E. L., and Aslin, R. N. (**1996b**). "Word segmentation: The role of distributional cues," J. Mem. Lang. **35**, 606–621.

Schreuder, R., and Baayen, R. H. (**1994**). "Prefix stripping re-revisited," J. Mem. Lang. **33**, 357–375.

Scott, D. (**1982**). "Duration as a cue to the perception of a phrase boundary," J. Acoust. Soc. Am. **71**, 996–1007.

Sluijter, A. M. C., van Heuven, V. J., and Pacilly, J. J. A. (**1997**). "Spectral balance as a cue in the perception of linguistic stress," J. Acoust. Soc. Am. **101**, 503–513.

Suomi, K., McQueen, J. M., and Cutler, A. (**1997**). "Vowel harmony and speech segmentation in Finnish," J. Mem. Lang. **36**, 422–444.

Thiessen, E. D., and Saffran, J. R. (**2003**). "When cues collide: Use of stress and statistical cues to word boundaries by 7- to 9-month-old infants," Dev. Psychol. **39**, 706–716.

Thiessen, E. D., and Saffran, J. R. (**2004**). "Spectral tilt as a cue to word segmentation in infancy and adulthood," Percept. Psychophys. **66**, 779–791.

Toro, J. M., Nespor, M., Mehler, J., and Bonatti, L. L. (**2008**). "Finding words and rules in a speech stream: Functional differences between vowels and consonants," Psychol. Sci. **19**, 137–144.

Toro-Soto, J. M., Rodríguez-Fornells, A., and Sebastián-Gallés, N. (**2007**). "Stress placement and word segmentation by Spanish speakers," Psicologica **28**, 167–176.

Tyler, M. D. (**2006**). "French listeners can use stress to segment words in an artificial language," in *Proceedings of the 11th Australasian International Conference on Speech Sci. & Tech.*, edited by P. Warren, and C. I. Watson (Australasian Speech Sci. and Technol. Assoc. Inc., Auckland, New Zealand), pp. 222–227.

Vaissière, J. (**1983**). "Language-independent prosodic features," in *Prosody: Models and Measurements*, edited by A. Cutler, and D. R. Ladd (Springer-Verlag, Hamburg), pp. 53–66.

van der Hulst, H. (**1999**). *Word Prosodic Systems in the Languages of Europe* (Mouton de Gruyter, Berlin).

van der Lugt, A. H. (**2001**). "The use of sequential probabilities in the segmentation of speech," Percept. Psychophys. **63**, 811–823.

Vroomen, J., Tuomainen, J., and de Gelder, B. (**1998**). "The roles of word stress and vowel harmony in speech segmentation," J. Mem. Lang. **38**, 133–149.

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

Woodrow, H. (**1909**). "A quantitative study of rhythm: The effect of variations in intensity, rate, and duration," Archives of Psychol. **14**, 1–66.