JOURNAL OF MEMORY AND LANGUAGE **38,** 133–149 (1998)
ARTICLE NO. ML972548

The Roles of Word Stress and Vowel Harmony in Speech Segmentation

Jean Vroomen

University of Tilburg, Tilburg, The Netherlands

Jyrki Tuomainen

University of Tilburg, Tilburg, The Netherlands; and Centre for Cognitive Neuroscience, University of Turku, Turku, Finland

and

Beatrice de Gelder

University of Tilburg, Tilburg, The Netherlands; and Universite Libre de Bruxelles, Belgium

Three experiments investigated the role of word stress and vowel harmony in speech segmentation. Finnish has fixed word stress on the initial syllable, and vowels from a front or back harmony set cannot co-occur within a word. In Experiment 1, we replicated the results of Suomi, McQueen, and Cutler (1997) showing that Finns use a mismatch in vowel harmony as a word boundary cue when the target-initial syllable is unstressed. Listeners found it easier to detect words such as *HYmy* in *PUhymy* (harmony mismatch) than in *PYhymy* (no harmony mismatch). In Experiment 2, words had stressed target-initial syllables (*HYmy* as in *pyHYmy* or *puHYmy*). Reaction times were now faster and the vowel harmony effect was greatly reduced. In Experiment 3, Finnish, Dutch, and French listeners learned to segment an artificial language. Performance was best when the phonological properties of the artificial language matched those of the native one. Finns profited, as in the previous experiments, from vowel harmony and word-initial stress; Dutch profited from word-initial stress, and French did not profit either from vowel-harmony or from word-initial stress. Vowel disharmony and word-initial stress are thus language-specific cues to word boundaries. © 1998 Academic Press

One of the major issues in spoken word recognition concerns the detection of word boundaries in continuous speech. The central problem is to understand how listeners segment the continuous speech signal into discrete words when there are no reliable acoustic cues that signal the beginnings of words. A number of alternative ideas have appeared in the literature that point toward a possible solution. A major division can be made between proposals that emphasize acoustic/phonetic cues and those that focus on lexical or contextual processes. In the former, word boundaries are located on the basis of local perceptual features such as the presence of glottal stops, laryngealized voicing, increased aspiration, or

The research was partly supported by a grant from the Human Frontier of Science Programme ‘‘Processing consequences of contrasting language phonologies.’’ The research of Jean Vroomen has been made possible by a fellowship of the Royal Netherlands Academy of Arts and Sciences. The research of Jyrki Tuomainen was financially supported by the Academy of Finland. Research was also partly supported by the Ministry of Education of the Belgian French-speaking Community, Concerted Research Action ‘‘Language processing in different modalities: Comparative approaches.’’ We thank Leo Vogten from the IPO, Eindhoven for help in preparing the stimuli of Experiment 3 and Juan Seguí for help in testing the French subjects. We also thank James McQueen, Arthur Samuel, Kari Suomi, James Sawusch, and two other anonymous reviewers for comments on an earlier version of the paper.

Address correspondence and reprint requests to Jean Vroomen, Department of Social Sciences, University of Tilburg, PO Box 90153, 5000 LE Tilburg, The Netherlands. E-mail: j.vroomen@kub.nl.

vowel lengthening (e.g., Lehiste, 1960; Nakatani & Schaffer, 1978). Proposals in the latter category use concepts such as the uniqueness point of the word, lexical competition, or ‘‘top-down’’ knowledge (e.g., Cole & Jakimik, 1980; Marslen-Wilson, 1984; McClelland & Elman, 1986; Norris, 1994).

In natural speech, both phonetic and lexical cues are present. For example, a word boundary can be signaled by the simultaneous presence of a long silence that precedes the word, word-final vowel lengthening (Umeda, 1975), or, in English, the aspiration of an initial stop (Nakatani & Dukes, 1977). In addition, segmentation is facilitated when the initial syllable of the word contains a full vowel (Cutler & Norris, 1988; Vroomen, van Zon, & de Gelder, 1996), when the word starts at the beginning of a syllable (Vroomen & de Gelder, 1997), or when few lexical competitors are present (McQueen, Norris, & Cutler, 1994; Norris, McQueen, & Cutler, 1995; Vroomen & de Gelder, 1995). Each of these factors on its own may not be sufficient, but they jointly point toward a likely word boundary.

Little research has focused on how listeners deal with multiple segmentation cues. Each of the previously mentioned cues has been studied in isolation, and it is unknown as yet what listeners do in the presence of multiple, possibly conflicting segmentation cues. One possibility is that the relative importance of one cue is weighted against others. If so, it is critical to study the respective weights of different cues and how they are combined. Another question is whether lexical and phonetic cues combine. In a similar vein, it is of interest to know whether segmentation cues have trading relations—like phonetic cues—so that one cue functions in the absence of another. One may also ask whether multiple segmentation cues work in an additive way, or, in the case of conflict, whether one cue is overruled by the others. A more complicated scenario is that, due to time constraints, some cues may only be effective in off-line tasks, but not in on-line speech segmentation.

In the present study, we explored some of these issues by examining word stress and vowel harmony as potential segmentation cues in Finnish. Finnish has front-back vowel harmony (Karlson, 1983). The Finnish vowels /u, a, o/ belong to the back harmony set, /y, æ, ø/ to the front harmony set, and /i, e/ are neutral. The main restriction in uncompounded Finnish words is that vowels from the front and the back harmony class cannot occur together, but vowels from the neutral class can be combined with both the front or back class vowels in any position in the word stem. Harmony propagates from left to right from the first vowel in the root to subsequent vowels in root and suffix. Vowels of suffixes are therefore subject to the harmony restriction. As an example, *kapula* (meaning *stick*) and *räjähdys* (*explosion*) are possible Finnish words because /a, u/ are from the back harmony class, and /æ, y/ are from the front harmony class (/æ/ is written as *ä*). The correctly suffixed forms of the words would be *kapulako* (meaning *a stick?*) and *räjähdyskö* (*an explosion?*, /ø/ is written as *ö*). But *kapyla* and *räjahdys* would be prohibited as Finnish words because their vowels are from opposing classes. A clash in vowel harmony (for example a front vowel followed by a back vowel, or vice versa) is in Finnish thus typically associated with a word boundary. (There are some exceptions to this rule such as *analyysi,* meaning *analysis.*)

The second potential segmentation cue we investigated is word stress or primary stress. Word stress is an abstract phonological property of a word that, under certain conditions, is phonetically realized so that the stressed syllable is more prominent or salient relative to the other syllables. Every word that belongs to a lexical category contains exactly one syllable that carries primary stress, while all other syllables are subordinated. In fluent speech, one can distinguish stressed syllables from other syllables because they tend to be louder, longer in duration, different in pitch, or—in English—their vowels are less centralized to schwa. In Finnish, the primary stressed syllable is always the initial syllable of the word. Accordingly, from a phonological point of view, word stress might be a reliable indicator of word boundaries. However, there are at least two potential problems with the use of word stress as a segmentation cue. The first is that word stress is an abstract property of the word that is not always acoustically realized in the speech signal. Listeners may thus be unable to perceive whether a syllable carries primary stress because there are no phonetic correlates. The second difficulty is that, even if stress is perceivable, it is not clear whether listeners actually use this information in on-line speech segmentation.

The potential use of stress as a cue to word boundaries was studied recently by Iivonen, Niemi, and Paananen (submitted), who tried to determine the extent to which fundamental frequency (F0) peaks in Finnish, English, and German coincide with word stress. They analysed TV and radio newscasts and counted how often a just noticeable F0 peak (defined as a difference in one semitone or more when compared with the neighboring syllable) matched a primary stressed syllable. One cannot expect a perfect correlation between F0 peaks and word stress because stress may not always be acoustically realized. In addition, not every F0 peak signals word stress, because it is well known that the F0 contour has other linguistic functions such as accentuation, signaling of emotions, or cueing of syntactic boundaries (see Cutler, Dahan, & van Donselaar, 1997 for a recent overview). These and other rhythmic phenomena such as the avoidance of stress clashes are likely to obscure the relation between word stress and its phonetic correlates. Nevertheless, Iivonen, Niemi, and Paananen found that the majority of Finnish F0 peaks, 73%, occurred on the primary stressed syllable, while only 42% of the German peaks and 59% of the English peaks represented word stress. Moreover, about 52% of the Finnish word-initial syllables had an F0 peak. Thus, this phonetic analysis suggests that F0 peaks are at least partly successful in signaling where primary stress is, and hence, where a word boundary is located in Finnish speech.

The actual use of word stress in speech segmentation has been contested by Cutler and colleagues (Fear, Cutler, & Butterfield, 1995). They have argued that it is not *word* stress but *metrical* stress that is used in on-line speech segmentation. Metrical stress is mainly based on whether a syllable's vowel is full or reduced. Fear et al. argued that word stress is not used in on-line speech segmentation because it is a syntagmatic property (a stressed syllable is stressed *relative* to the others). In contrast, metrical stress is a paradigmatic property which can be perceived in absolute terms. The judgement about whether or not vowel quality is reduced can be made immediately, but relational judgements about whether one syllable is more prominent than the other are thought to be time consuming. Hence, the argument is that word stress can only be determined post-lexically, which led Fear et al. to infer that word stress is unlikely to be used in on-line word recognition.

In our view, the role of word stress in speech segmentation is still a matter of debate because so far little is known about the role of word stress in different languages. Moreover, the presumption that word stress can only be determined post-lexically may be wrong. It seems possible that a stressed syllable can be perceived as stressed without reference to neighboring syllables, for example on the basis of characteristic F0 transitions within the syllable, a long duration, or an increased intensity (of higher harmonics). In addition, a stressed syllable in continuous speech may stand out relative to the previous syllable. Given that almost all Finnish words are multisyllabic with unstressed final syllables, stressed syllables are usually preceded by the unstressed word-final syllable of the preceding word. For these reasons, stressed syllables may be perceived as stressed even though the word to which they belong is not yet recognized. There is therefore no strong a priori reason to rule out word stress as a segmentation cue.

To investigate the combined roles of word stress and vowel harmony in speech segmentation, we conducted a study in which both factors were varied. Experiment 1 was a replication of Suomi et al. (1997, Experiments 1 and 4) in which word boundaries did not have a stress cue. Listeners had to detect words such as *HYmy* (the stressed syllable is denoted with capital letters) in *PUhymy* (harmony clash between prefix and target word; no stress cue on the first syllable of the embedded word) or *PYhymy* (no harmony clash; no stress). This replication was conducted first in order to have a basis for later comparisons. It also allowed us to check whether we had artifacts in items, participants, equipment, or procedures that might explain any deviant results. Experiment 2 was similar to the previous one, except that target words now contained a stress cue such as *HYmy* in *puHYmy* or *pyHYmy.* In Experiment 3, we used an artificial learning task in which Finnish, French, and Dutch speakers had to segment an artifical language into ''words.'' This allowed us test the generality of our findings across tasks and to examine the extent to which vowel harmony and word stress are language-specific cues to word boundaries.

EXPERIMENT 1

The task of the listeners was to detect bisyllabic CVCV words (C = consonant, V = vowel) which were preceded by a CV prefix. The vowel of the prefix was either harmonious with the vowels of the embedded target word or not. The CVCVCV string always had primary stress on the prefix so that the embedded target word had no stress cue. Suomi et al. (1997) found that listeners use vowel disharmony as a cue for speech segmentation. Thus, *HYmy* was easier to detect in *PUhymy* than in *PYhymy.*

Method

Participants. Twenty native Finnish speakers took part in the experiment. They were students from an introductory psychology class or staff members from the Centre for Cognitive Neuroscience of the University of Turku. All reported normal hearing. Equal numbers received both versions of the test.

Materials. The same experimental items were used as in Suomi et al. (1997). They were spoken by JT and recorded anew. Thirty CVCV target words were employed. Half contained vowels from the back harmony class, and half from the front harmony class. All words were monomorphemic nouns or adjectives in their uninflected form. Two alternative CV prefixes were used to create a nonword that contained the embedded word at its end. For each item, one prefix contained a vowel that belonged to the same harmony class as the vowels of the target, and one had a vowel from the opposite class. All items were pronounced with lexical stress on the prefix. For example, the word *PAlo* (fire) had as prefixes *ku* and *ky,* and was thus pronounced as *KUpalo* or *KYpalo.* This produced 60 trisyllabic items, none of which contained any other word besides the intended one. The target-bearing items are listed in the Appendix.

Another 60 trisyllabic CVCVCV filler items were created that did not contain an embedded word. In half of them the two final vowels were from the back harmonic class, and in the other half they were from the front class. Within both sets, half of the items had a first syllable that was harmonious with the rest, while in the other half the first vowel was disharmonious with the rest. All fillers had, like the experimental items, stress on the initial syllable.

The materials were recorded in a sound-treated room on DAT tape. The items were then digitized at 22.05 kHz with 16 bits precision, and the onset and offset of the embedded words were determined with a speech editor under auditory and visual control. The items were played to participants directly from the hard disk of a PC.

Design and procedure. Two lists were constructed, so that a participant heard each embedded target word only once. The type of context was counterbalanced over the lists. The position of fillers and each member of an experimental item pair was the same in the two lists. A short practice session of 16 trials preceded the experiment.

Participants were tested individually in a quiet room. All items were presented over a loudspeaker with an inter-trial interval of 4.5 s. Participants were instructed that they would hear a nonsense item which sometimes contained a finally embedded real word. They were asked to press a button with their preferred index finger as soon as they heard a real word, and then to say the word aloud. The vocal response was checked by the experimenter to determine whether the intended word had been detected correctly.

Results and Discussion

Unless stated otherwise, all analyses were done in exactly the same way as by Suomi et al. (1997). Reaction times (RT) were measured from the offset of the word, and vocal responses that did not correspond to the intended word (0%) and outlying responses (4%) were treated as errors and discarded from the RT analyses. Outlying responses were defined as RTs slower than 2000 ms as measured from target offset. It should be noted that Suomi et al. used the same upper cut-off criterion, but they also discarded RTs faster than 150 ms. In our Experiment 1, no response was faster than this criterion. However, in our Experiment 2 responses were much faster, and in that case it would not have been correct to treat RTs faster than 150 ms as ‘‘outliers.’’ For consistency across our experiments, we therefore discarded only responses longer than 2000 ms. The false alarm rate (i.e., a key response on a filler item) was 2.1%. Inspection of individual items and participants showed that no item was missed by more than 50% of the subjects and no participant made more than 50% errors. No participant or item was therefore excluded. The mean RTs and error rates are presented in the top panel of Table 1.

Analyses of variance (ANOVA) was performed with subjects (*F1*) and items (*F2*) as repeated measures. In the subject analyses, harmony class of the target word (back or front vowel) and prefix type (harmonious or disharmonious) were within-subjects variables, and in the item analyses, harmony class of the target word was a between-items factor, and prefix type was a within-items factor. A 2 × 2 ANOVA showed that, in the subject analysis, target words with a disharmonious prefix (*HYmy* in *PUhymy*) were detected 112 ms faster than targets with a harmonious prefix (*HYmy* in *PYhymy*), $F1(1,19) = 36.15$, $p < .001$; but this effect was only marginally significant in the item analysis, $F2(1,28) = 2.85$, $p = .10$. There was no overall difference between targets with vowels from the front or back harmony class, $F(1,13) = 1.48$, ns; $F2 < 1$, and only in the subject analysis did the harmony effect interact with the harmony class of the target, $F1(1,19) = 4.86$, $p < .05$; $F2 < 1$. Inspection of Table 1 shows that the harmony effect was larger for targets with vowels from the front harmony class (203 ms) than for targets from the back harmony class (91 ms). Separate tests showed that the harmony effect for targets from the back harmony class was significant by subjects only, $F1(1,19) = 5.60$, $p < .05$, $F2 < 1$. For targets from the front harmony class, the harmony effect was significant by subjects, $F1(1,19) = 40.18$, $p < .001$, and marginally significant by items, $F2(1,14) = 3.90$, $p = .06$.

The RTs of our Experiment 1 were very similar to those of Suomi et al. (1997), which are presented in the bottom of Table 1. They found that disharmonious items were detected faster than harmonious items (161 ms on average; we obtained a 147 ms effect), and they also obtained an interaction showing that the effect was reliable for targets with front vowels (218 ms; we obtained a 203 ms effect), but not for targets with back vowels (103 ms; we obtained a 91 ms effect). Also, as in the present experiment, their item analyses were less significant (smaller *F* values and *p* values less significant) than the subject analyses. This is mainly due to the fact that there are large differences among items which are not controlled for frequency of occurrence, familiarity, imageability, or onset phoneme. Finally, the average RT in Suomi et al.’s study was somewhat faster than in our experiment (731 ms versus 807 ms). In absolute terms, though, RTs were slow in both experiments if one considers that they were measured from word offset.

Analysis of the error rates showed no trend

TABLE 1

Mean Reaction Time (in Milliseconds) and Error Rate (in Parentheses) in Experiment 1 and Suomi et al. (1997)

	RT from target offset		RT from target onset	
	Target		Target	
Experiment 1 context	Back	Front	Back	Front
Harmonious	870 (12%)	891 (15%)	1228 (16%)	1206 (22%)
Disharmonious	779 (9%)	688 (9%)	1122 (13%)	1042 (10%)
Suomi et al. (1997)				
Harmonious	802 (9%)	822 (10%)		
Disharmonious	699 (5%)	604 (4%)		

toward a speed–accuracy trade-off. The ANOVA on the errors by subjects showed that more targets were missed when the prefix was harmonious than when it was disharmonious (13% vs 9%), $F1(1,19) = 4.93$, $p < .05$, but this difference was not significant in the item analysis, $F2 < 1$. No other main effect or interaction was significant (all F's < 1). This error pattern is again very similar to that of Suomi et al. (1997). In their Experiment 1, they found a significant main effect of prefix in the same direction as ours, but no other effects were significant.

In the following analyses, duration of the target was taken into account in order to check whether the RT effects were confounded by acoustic differences of the target words. The average duration of target words was 387 ms in harmonious strings and 376 ms in disharmonious strings (the items of Suomi et al. (1997) had similar durations of 374 ms and 393 ms in harmonious and disharmonious strings, respectively). Targets in harmonious strings were thus 11 ms longer than those in disharmonious strings, a difference that was significant in a t test $t(29) = 3.53$, $p < .001$. However, the difference in duration is in the wrong direction to account for the harmony effect, because when RT is measured from word offset, faster responses are usually found with longer words. Moreover, there was no correlation between the duration of the word and mean RTs or error rates in harmonious and disharmonious strings (all r's around $-.06$, and all p's $> .10$), and there was also no correlation between the size of the harmony effect and the difference in duration of the targets, $r(29) = .09$, $p = .62$. Separate correlational analyses for back and front words did not change this pattern (again all r's $< .10$ and all p's > 10). As in Suomi et al., it thus seems that differences in durations of the targets cannot account for the harmony effect.

As a further control for the duration of the items, we measured RTs from word onset (see Table 1). In this analysis, we again discarded RTs longer than 2000 ms, this time measured from word onset. This follows Suomi et al. (1997), even though it is debatable whether the same cut-off criterion of 2000 ms can be justified because more RTs than in the previous analyses had to be discarded (8% versus 4%). There was a harmony effect of 135 ms which was significant by subjects only, $F1(1,19) = 59.28$, $p < .001$, $F2(1,28) = 2.32$, $p = .13$. The interaction with harmony class of the target was not significant, $F(1,19) = 1.11$, ns; $F2 < 1$. Pairwise comparison showed that the harmony effect was significant in the subject analysis for targets with back vowels, $F1(1,19) = 7.97$, $p < .02$, and for targets with front vowels, $F1(1,19) = 36.05$, $p < .001$, but the effects were not significant in the item analysis (both p's $> .10$). Thus, the results of the item analyses in which RT was measured from word onset were somewhat weaker than those in which RT was measured from word offset, but this is understandable because more RTs were discarded that passed the time-out criterion. The results are again similar to the results of Experiment 1 of Suomi et al. (1997) in which there was also no significant interaction in the item analysis. No comparison can be made with their Experiment 4, because these analyses were not reported.

We also performed a new analysis on the error rates because more responses passed the 2000 ms time-out criterion. The subject analysis now showed that more errors were made with a harmonious prefix (19%) than with a disharmonious prefix (11%), $F1(1,19) = 8.72$, $p < .001$, but this difference did not reach significance in the item analysis, $F2(1,28) = 1.60$, NS). There was also a significant interaction in the subject analysis between prefix type and harmony class of the target, $F1(1,19) = 4.38$, $p < .05$; $F2 < 1$, showing that the difference between a harmonious and a disharmonious prefix was bigger in targets with vowels from the front harmony class (12% difference) than in targets with vowels from the back harmony class (3%).

All in all, we closely replicated the data of Suomi et al. (1997). There was an effect of vowel harmony which was stronger in words from the front harmony class than words from the back harmony class. This convergence allows us to continue our investigation, because we can now more safely account for differences that we may obtain in our next experiment.

EXPERIMENT 2

In Experiment 2 we investigated whether word stress plays a role in speech segmentation and whether the vowel harmony effect remains the same when the onset of the target is signaled by a stress cue. Suomi et al. (1997) argued that Finnish listeners do not use word stress in speech segmentation. They came to that conclusion because they could not find a difference between target words that did or did not have a stress cue (their Experiment 5). Their target words with a stress cue, such as *HYmy,* were spliced with a waveform editor from the beginning of a pseudoword, *HYmypu;* their targets without a stress cue, *hymy,* were spliced from the end of a pseudoword, *PUhymy.* However, this procedure allows a potential confound, because, in our experience, several prosodic and coarticulatory effects differently affect words spliced from the beginning or the end of a string. For example, the pitch of a word spoken in isolation usually ends within a more or less fixed region (This is similar to ‘t Hart, Collier, & Cohen, 1990 where sentence intonation is modeled by using a fixed end point of 75 Hz). The word *hymy* spliced from *HYmypu* may therefore sound strange because its pitch is at the end not back to the baseline. In contrast, the pitch in *hymy* spliced from *PUhymy* should sound normal in this respect. (This difference may help to explain why responses to items with a stress cue in Suomi et al.’s Experiment 5 were actually *slower* than responses to items without a stress cue.) Also, splicing *hymy* from *PUhymy* changes the relative prominence relations of the syllables in the target word because *hy* now becomes the most salient syllable, but this is not the case in *HYmy* spliced from *HYmypu.* Finally, and probably most important, it is questionable whether one can investigate the role of stress in speech segmentation if the target is presented in isolation (as in Suomi et al.’s Experiment 5). In that case, listeners do not need to segment the speech string because the signal is already parsed. Splicing may therefore not be an appropriate control to investigate the role of word stress in speech segmentation.

In our Experiment 2, instead of splicing, we rerecorded the same items in the same context, but the speaker now stressed the onset of the embedded word as would be done in natural speech. Thus, *HYmy* had to be detected in *puHYmy* (harmony clash, stress cue present) or *pyHYmy* (no harmony clash, stress cue present). If Finnish listeners use stress cues in word segmentation, then items with a stress cue should be easier to detect than those without. At this stage, no prediction can be made about the role of vowel harmony. According to Suomi et al. (1997), vowel harmony should be as effective as in non-stressed items. How-

TABLE 2

Mean Reaction (in Milliseconds) and Error Rate (in Parentheses) in Experiment 2

	RT from target offset		RT from target onset	
	Target		Target	
Context	Back	Front	Back	Front
Harmonious	270 (5%)	285 (9%)	696 (5%)	712 (9%)
Disharmonious	286 (5%)	270 (9%)	702 (5%)	678 (9%)

ever, an interaction between stress and vowel harmony would contradict this conclusion and would shed light on the relative contribution of vowel harmony and stress.

Method

Participants. Twenty students participated in the experiment. None had taken part in the previous experiment.

Materials. The same speaker, JT, had recorded the items of Experiment 1 and 2 at the same time. In Experiment 2 , items had stress on the first syllable of the embedded target word. The filler items were also recorded anew so that their stress pattern matched that of the experimental trials (i.e., stress on the second syllable of a trisyllabic string). All other experimental details were the same as in Experiment 1.

Results and Discussion

The RTs measured from word offset and error rates are presented in Table 2. There were no outliers (RTs equal or greater than 2000 ms), and analysis of the vocal responses showed that each target word was perceived as intended. The false alarm rate was 1.5%, which is not significantly different from the 2.1% in Experiment 1, $F(1,38) < 1$. The same analyses on RTs and error rates were performed as in Experiment 1. In the 2 × 2 ANOVA on the RTs, there was no effect of harmony (both F's < 1), no difference between targets with front and back vowels (both F's < 1), and no significant interaction (all p's $> .10$).

In the ANOVA on error rates there was again no difference between harmonious or disharmonious items (both F's < 1). There was a trend for targets with front vowels to be missed more often than targets with back vowels, $F1(1,19) = 4.16$, $p = .056$; $F2(1,28) = 5.03$, $p < .05$, but this did not interact with the harmony effect (both F's < 1).

The durations of the targets were 427 ms and 416 ms in the harmonious and disharmonious context respectively, $t(29) = 3.97$, $p < .001$. As in the previous experiment, all correlations between the overall RT and duration of the targets were small and non-significant.

When RTs were measured from word onset, there was in the 2 × 2 ANOVA a small harmony effect in the subject analysis, $F1(1,19) = 5.15$, $p < .05$, but it was not significant in the item analysis, $F2 < 1$. There was also a trend for an interaction, but again it was not significant, $F1(1,19) = 3.46$, $p = .08$; $F2(1,28) = 3.03$, $p = .10$.

The crucial analysis is the comparison between Experiment 1 and 2, because that will show whether stress had an effect and whether it changed the harmony effect. An ANOVA was conducted on the RTs in which Experiment was a between-subjects and a within-items factor. When RTs were measured from word offset, there was a main effect of Experiment because RTs were much faster in Experiment 2 than in Experiment 1, $F1(1,38) = 66.38$, $p < .001$; $F2(1,28) = 984.56$, $p < .001$. There was also an interaction between Experiment and harmonious/disharmonious prefix showing that the harmony effect was present in Experiment 1 (147 ms), but not in Experiment 2 (0 ms), $F1(1,32) = 31.57$, $p < .001$; $F2(1,28) = 4.14$, $p = .05$. When RTs were measured from word onset, there was again a main effect of Experiment, $F1(1,38) = 59.64$, $p < .001$; $F2(1,28) = 731.13$, $p < .001$. The interaction between Experiment and the harmony effect was significant by subjects only, $F1(1,38) = 41.37$, $p < .001$; $F2(2,28) = 2.28$, $p = .14$.

The same between-experiment analyses were performed on the error rates. In the item analysis more errors were made in Experiment 1 than in Experiment 2, $F2(1,28) = 4.20$, $p = .05$, but this was not significant in the subject analysis ($p > .10$). The interaction between Experiment and harmony of the prefix was not significant in the error analysis ($p > .10$).

To summarize, we found that words with a stress cue had a much faster RT and a much smaller harmony effect than words without a stress cue. This contradicts the conclusion of Suomi et al. (1997), who argued that word stress does not play a role in the recognition of Finnish words. In stark contrast with their conclusion, our results show that word stress plays an important role in the segmentation of Finnish speech. Finnish listeners take stressed syllables as a potential word onset, and this explains why, for example, *hymy* is so much faster to detect in *puHYmy* than in *PUhymy.* Moreover, when words are stressed, stress is such a strong cue that there is no room for a contribution of vowel harmony. This thus suggests that the contribution of word stress is more important than that of vowel harmony. In our next experiment, we tried to confirm this conclusion with a different task.

EXPERIMENT 3

In Experiment 3, we adopted an entirely different paradigm from the word spotting task. If the results of this new task converge with those of the word spotting experiments, it would considerably strengthen our conclusion about the role of vowel harmony and word stress. It would then become more likely that the observed pattern is not a specific feature of the word spotting task, but a genuine aspect of speech processing.

In our new task, listeners were confronted with a completely unknown artificial ‘language’ that none had ever heard before. The language was made up of ‘words’ that were concatenated in random order into a long continuous string of synthesized speech with no pauses between the words. The task of the listener was to discover the words of which the language was made up (see Saffran, Newport, & Aslin, 1996 for previous use of this task). In different conditions, words contained either harmonious or disharmonious vowels, and the word’s initial syllable was either stressed or not. The results of Experiments 1 and 2 lead us to predict that in the absence of a stress cue, Finns should find harmonious words easier to segment than disharmonious words. However, when the initial syllable is stressed, Finns should find the task much easier and there should be no difference between harmonious and disharmonious words.

The above prediction is based on the assumption that listeners bring their native segmentation routine to the task of learning an artificial language. We thus assume that adult listeners do not start from zero, but rather that they give weight to those speech cues which have significance in their native language. This notion is in line with the results of Cutler, Mehler, Norris, and Segui (1986). They found that French monolinguals use their native segmentation routine when listening to an unknown foreign language, which in their study was English. This led Cutler et al. to conclude that monolinguals have a language-specific segmentation routine which they cannot switch off when listening to a foreign language. Our concern in the present experiment, though, was whether listeners would rely on their native segmentation routine when listening to artificial synthesized language which lacks the naturalness and richness of real speech.

To determine whether listeners apply their native segmentation routine when performing the learning task, we presented the same materials to listeners from different language backgrounds. For the present comparison, French is maximally different from Finnish because French does not have vowel harmony, and stress in French polysyllabic words is never on the initial syllable but always on the last full vowel of content words (Dell & Vergnaud, 1984). If the task reflects properties of the native segmentation routine, then French listeners should not be influenced by whether words are harmonious or disharmonious. Also, word-initial stress should not be helpful because that conflicts with the French stress pattern.

An intermediate case between Finnish and French is Dutch. Dutch, like French, has no vowel harmony. We therefore expected Dutch listeners not to be sensitive to vowel harmony. The position of the stressed syllable is, unlike Finnish and French, variable in Dutch. According to Kager (1989), the penultimate position receives primary stress as default, but a count in the Dutch CELEX lexicon showed that most multi-syllabic words have stress on the initial syllable. Of all two-, three-, and four-syllabic words with a frequency of occurrence higher than or equal to one, 56% of the tokens had lexical stress on the first syllable (15,357 entries out of 27,020 selected words). For tri-syllabic words, as were used in the present experiment, 53% had stress on the initial syllable (6220 words out of all 11,646 trisyllabic words), 32% (or 3788 words) had stress on the penultimate syllable, and 14% (1638 words) had stress on the final syllable. Taking these statistical facts into account, stressed syllables are likely to be a word onset in Dutch, and Dutch listeners may therefore profit from a stress cue on the word-initial syllable.

Method

Participants. Three different native-language groups were tested: Finnish, Dutch, and French. There were 43 Finns, 53 Dutch, and 44 French. Participants were recruited from introductory Psychology classes or, occasionally, were staff members. The Finns were recruited from the Centre for Cognitive Neuroscience and the University Hospital of Turku, the Dutch were recruited from the University of Tilburg, and the French were recruited from the Université René-Descartes, Paris. Each participant heard only one out of four different artificial languages. Participants received course credit or a small amount of money.

Materials. For the learning phase, an artificial language was constructed consisting of four consonants (/v/, /m/, /t/, and /k/) and six vowels (/o/, /u/, /a/, /y/, /ɛ/, and /œ/) that made up 15 different CV syllables. The syllables were combined so as to create two separate lexicons, a harmonious and a disharmonious one, each consisting of six trisyllabic words. The words in the harmonious lexicon had vowels belonging either to the front harmony set (/y/, /œ/ and /ɛ/) or the back harmony set (/u/, /o/ and /a/). The back harmony words were /vomuvu/, /tokuvo/, and /motamu/; the front harmony words were /mymɛty/, /vykɛvɛ/, and /tykɛty/. The words in the disharmonious lexicon were created by replacing one or two vowels of the harmonious words so that /o/ became /œ/, /ɛ/ became /a/ and /u/ became /y/. This resulted in the words /vœmyvu/, /tokuvœ/, /motamy/, /mumɛty/, /vykavɛ/, and /tykaty/. None of the items was, in any obvious sense, similar to a real Finnish, French, or Dutch word.

For both lexicons (harmonious and disharmonious), two versions with a different stress pattern were created. In the no-stress versions, all the words' syllables had equal stress, whereas in the stress-initial versions, the first syllable of each word received a pitch accent. This resulted in four experimental versions. Each version consisted of 150 tokens of the six words (total of 900 words, 2700 syllables). The words were concatenated in random order without spaces into a text file with the restriction that the same word could not occur twice in a row. The four versions had the same random order. The text file was split into 5 blocks of equal length, and each file was then input to the Spengi text-to-speech synthesizer at the Institute for Perception Research (IPO) in Eindhoven, which is based on Dutch diphone synthesis. The synthesizer speech rate was adjusted to a natural speech rate of approximately 275 syllables per minute. The phoneme durations were kept constant in all versions. In the no-stress version, the fundamental frequency was kept monotonous at 120 Hz throughout the whole string. In the stress-initial version, stress on the initial syllable was acoustically realized by using a pitch accent. The synthesis parameters for the F0 were set to its default values. The F0 linearly increased on the first syllable from 120 Hz to 170 Hz, and then gradually decreased over the next two syllables back to baseline.[1] The synthesizer output was saved on an audio file (AIFF format, 16 bit precision, 16 kHz sampling rate), and each file was then recorded directly from a Silicon Graphics Iris Indigo workstation on a DAT tape.

For the test phase, three nonwords foils (for the harmonious version: /vutato/, /kutavo/, /vytymϵ/; for the disharmonious version, /vytyto/, /kutavϵ/, /vytamϵ/) and three part-word foils (for the harmonious version: /vomuto/, /kϵmϵty/, vykϵmy/; for the disharmonious version: /vœmuto/, /kumϵty/, /vykamy/) were created with the same technique and apparatus as the learning stimuli. Nonword foils contained the same syllables as were presented during the learning phase, but their order was not identical with any of the words. Part-word foils shared the initial or final two syllables with one of the real words. For the no-stress versions, foils did not have a stressed syllable; for the stress-initial versions, foils had the same pitch accent as the words.

Apparatus. All tapes were played back in a quiet room using a DAT-recorder and a high-quality loudspeaker. Participants were seated around a table and the speaker was located in front of the subjects at the distance of about 2.5 m.

Design and procedure. Participants were tested in groups of two to eight. As far as possible, equal number of listeners received one of the four versions. They were instructed to listen to the nonsense language and were told that the language consists of 'words' with no meaning or syntax. Their task was to figure out what the words were. They were given no information about the length or the number of words. During the learning phase, they were asked to listen to five blocks of 2 min each. There was a 5 s pause between the blocks. Participants were told that at the end a word recognition test was to be administered. The test was a two-alternative forced-choice task. Each test trial started with a tone, followed by a pair of trisyllabic strings separated by 500 ms of silence. One of the strings was a word of the artificial language, the other was one of the foils. Participants were asked to indicate whether the word came in first or second position by circling a ''1'' or ''2'' on a prepared answer sheet. They were told to guess if unsure and they were given 4 s for this. The complete test consisted of 36 trials (six words exhaustively paired with the six foils) with a short break in the middle. Four practice trials were given to acquaint participants to the structure of the test.

Results

The percentage of correctly recognized words in the two-alternative forced-choice test was computed for each listener. Table 3 presents the means across subjects. Simple *t* tests showed that performance in each of the twelve cells was significantly above chance (all *p*'s $< .05$ with a chance level of .5). An overall ANOVA with native language, stress, and vowel harmony as between-subjects factors showed that there was a main effect of language, $F(2,134) = 14.87$, $p < .001$, a main effect of stress, $F(1,134) = 20.65$, $p < .001$, and a significant interaction between language and stress, $F(2,134) = 3.33$, $p < .05$. The effect of vowel harmony and all other interactions with vowel harmony were not significant. Separate ANOVAs for each language group showed that Finns, $F(1,39) = 19.86$, $p < .001$, and Dutch, $F(1,49) = 10.83$, $p < .002$, profited from stress, but the French did

[1] Saffran, Newport, and Aslin (1996) used lengthening of the vowel as a cue for lexical stress. With English, they did not find an improvement when the word-initial vowel was lengthened. We conjecture that, at least for Finnish and Dutch, pitch accent is a better realization of stress for word-initial syllables than vowel lengthening (see, for example, 't Hart, Collier, & Cohen, 1990.)

TABLE 3

Mean Percentage of Correctly Identified Words by Finnish, Dutch, and French Listeners in Experiment 3

Vowel harmony	Finnish		Dutch		French	
	No stress	Stress	No stress	Stress	No stress	Stress
Harmonious	72%	86%	65%	79%	58%	58%
Disharmonious	64%	85%	64%	75%	62%	67%

not ($F < 1$). Inspection of Table 3 shows that in the Finnish group there was a trend toward an interaction between stress and vowel harmony in the predicted direction, but this trend was statistically not significant, $F(1,39) = 1.11$, $p = .299$. Despite the lack of a significant interaction, separate t tests were conducted because the between-subjects design is statistically rather conservative. However, t tests in which the harmony effect is tested should be interpreted with caution, because the harmony effect or its interaction was not significant in the overall ANOVA.

Finnish listeners. In the no-stress condition, harmonious words were recognized better than disharmonious words. A t test (one-tailed) for independent samples showed that the 9% difference was significant, $t(22) = 2.21$, $p < .02$. In order to ensure that this effect did not depend on just a few listeners performing extremely well (or poorly), we conducted another by-subjects analysis by determining whether each listener's performance was better than expected by chance. According to a binomial test (with $p < .05$), performance at or above 66% in a 36-item test is significantly better than chance. For each condition, then, the number of participants performing above this level was determined, and a chi-square test was used to test whether there was a statistically reliable difference between conditions. In the no-stress disharmonious condition, 5 out of 12 (41%) listeners performed above chance, and in the harmonious condition 11 out 12 listeners (91%). According to a chi-square test, this difference is significant, $\chi^2_{(1)} = 6.75$, $p < .01$. Thus, more Finnish listeners performed above chance with harmonious items than with disharmonious items.

In the stress-initial condition, there was no difference between harmonious and disharmonious items, $t(17) = -.13$, NS. With harmonious items, eight out of nine participants (89%) performed better than chance, and with disharmonious items 9 out of 10 participants (90%), $\chi^2_{(1)} < 1$. Moreover, average performance in the stress-initial conditions was much better than in the no-stress conditions. Overall performance increased from 69% in the no-stress conditions to 86% in the stress-initial conditions, an increase of 16%. Simple t tests showed that the improvement was significant for harmonious, $t(20) = 2.47$, $p < .02$ and disharmonious items, $t(19) = 3.80$, $p < .001$.

Dutch listeners. Dutch participants did not show a difference in the no-stress condition between harmonious and disharmonious items, $t(24) = -.24$, NS. In both conditions, 7 out of 13 participants (54%) performed above chance (no testing required). With stress-initial words, there was also no difference between the harmonious and disharmonious items, $t(25) = -.63$, NS. With stress-initial harmonious items, 10 out of 13 participants (77%) performed better than chance, and with stress-initial disharmonious items 11 out of 14 participants (78%), $\chi^2_{(1)} < 1$. The Dutch improved when words had stress on the initial syllable (on average 65% for no-stress items versus 77% for stress-initial items, an increase of 12%). The improvement was significant both for harmonious, $t(24) = 2.57$, $p < .01$, and disharmonious items $t(25) = 2.08$, $p < .03$.

French listeners. There was no difference between harmonious and disharmonious no-stress items, $t(21) = .67$, NS. With harmonious items, 3 out of 10 participants (30%) performed above chance, and with disharmonious items 7 out of 13 participants (53%), $\chi^2_{(1)} = 1.30$, NS. With stress-initial words, there was also no difference between harmonious and disharmonious items, $t(19) = 1.41$, p = NS. With stress-initial harmonious items, 5 out of 11 participants (45%) performed above chance, with stress-initial disharmonious items 3 out of 10 participants (30%), $\chi^2_{(1)} < 1$. Neither with harmonious, $t(18) = .09$, NS, nor with disharmonious items, $t(22) = -.82$, NS, was there a difference between the no-stress and stress-initial items. French listeners thus profited neither from vowel harmony nor from word-initial stress.

Between-Language Comparisons

Finnish versus Dutch. From all pairwise comparisons between Dutch and Finns, only one was marginally significant showing that the no-stress harmonious items were recognized better by the Finns than the Dutch, $t(23) = 1.83$, $p = .08$, $\chi^2_{(1)} = 4.42$, $p < .05$. All other comparisons did not reach significance (all p's $> .10$).

Finnish versus French. Finns did not differ from the French with disharmonious no-stress items, $t(23) = .49$, NS; $\chi^2_{(1)} < 1$, but the harmonious no-stress items were recognized better by the Finns than the French, $t(20) = 3.21$, $p < .005$; $\chi^2_{(1)} = 8.96$, $p < .01$. With stress-initial items, Finns performed better than French with harmonious, $t(18) = 4.31$, $p < .001$; $\chi^2_{(1)} < 4.10$, $p < .05$, and disharmonious items, $t(18) = 2.62$, $p < .02$; $\chi^2_{(1)} = 7.50$, $p < .01$.

Dutch versus French. There was no difference between Dutch and French with harmonious and disharmonious no-stress items (all p's > 10). However, stress-initial harmonious items were recognized better by the Dutch than by the French, $t(21) = 3.58$, $p < .002$, $\chi^2_{(1)} = 5.06$, $p < .05$. The better performance of the Dutch with stress-initial disharmonious items failed to reach statistical significance, $t(23) = 1.21$, $p = .24$, $\chi^2_{(1)} = 2.93$, $p < .10$.

Discussion

The results show that Finns and Dutch profit from a stress cue on the word-initial syllable, but the French do not. This result is in line with the phonological properties of the languages. Finnish words always have word-initial stress, in Dutch the majority of words have word-initial stress, but in French no words have initial stress. Moreover, the vowel harmony effect was only observed with Finnish listeners in words without a stress cue. The Finnish results of the learning task are therefore in close correspondence with the word-spotting experiments. Again they show that Finns use stress and vowel harmony as cues to word boundaries, and that the presence of a stress cue greatly reduces the contribution of vowel harmony.

Experiment 3 shows that the artificial learning task has the potential to provide insights into language-specific aspects of speech processing. Finnish, Dutch, and French listeners were helped when the phonological properties of the artificial language matched those of their native language. It thus appears that the task is sensitive to the cues that listeners use when segmenting their native language. The learning task is therefore a promising tool for further research because it allows careful control over the phonological properties of the artificial language and the amount of exposure listeners receive.

GENERAL DISCUSSION

In three experiments, we observed that Finns use, in an interdependent way, vowel disharmony and word stress as cues to word boundaries. In a word spotting task, vowel disharmony was used when the word-initial syllable was unstressed, but the effect was greatly reduced when there was a stress cue on the word-initial syllable. The same pattern was obtained in a learning task: Finns found harmonious words without a stress cue easier to segment than comparable disharmonious words, but the presence of a stress cue improved performance and the difference between harmonious and disharmonious words disappeared. These results are in direct contrast with the conclusion of Suomi et al. (1997), who argued that ‘‘word stress may not play an important role in recognition of Finnish speech.’’ They further stated that ‘‘It is very unlikely that the harmony mismatch effects emerged because of the absence of canonical stress cues.’’ It now seems clear that this conclusion cannot be maintained. In fact, the opposite is the case: Stress is the strongest cue, and it greatly reduces the effect of vowel harmony. The results of Suomi et al. can therefore not be generalized to normal fluent speech where stressed syllables are often signaled by F0 peaks or other stress cues (see Iivonen et al., submitted).

Why does prominence reduce the contribution of vowel disharmony? Even though a stress cue may be more important than vowel disharmony, it does not mean that the role of vowel disharmony should be diminished. In fact, in perception it seems to be more the rule than the exception that cues are only partly valid. So the question is why vowel disharmony is not used in conjunction with stress.

One possibility is that listeners do not rely heavily on vowel disharmony because many words are missed that do not have a vowel disharmony cue. It may therefore be critical to have an estimate of the success rate of an algorithm that detects vowel disharmonies. We addressed this issue by running a simple statistic on two samples of text (one 654 words long, the other 601) taken from a 1996 issue of a monthly supplement to the Finnish main newspaper (Helsingin Sanomat). Our ‘‘vowel disharmony’’ algorithm assumed a word boundary between two adjacent syllables any time their vowels changed from either back to front or from front to back. As an example, the algorithm would correctly detect the word boundary between *syövät jonkun* (eat someone) because the vowels across the words change from front to back. Using this criterion, the algorithm correctly detected 19% of the word boundaries in the first text, and 17.5% in the second one. The false alarm rate was 2.1 and 2.5% respectively, mainly stemming from compound words that did not have vowel harmony (e.g., *polkupyörä*, meaning *bicycle*). The reason for this rather low hit rate is that in many cases adjacent words are from the same harmony class, because, among other factors, there are more words from the back harmony class than words from the front harmony class. Moreover, many Finnish words contain neutral vowels that can occur in any position within a word. Changes from neutral to back or neutral to front, or vice versa are therefore not informative about the presence of a word boundary. The situation worsens if one takes into account that both we and Suomi et al. (1997) observed that the harmony effect was only significant in targets with front vowels, but not for targets with back vowels. Finnish listeners were thus more sensitive to a back to front than to a front to back change (for a possible explanation of this asymmetry, see Suomi et al.). If only the back to front change is counted, then the success rate of the vowel harmony algorithm further dropped to only 6.4% in text 1 and 5.8% in text 2. These statistical properties thus show that the a priori success rate of a vowel disharmony algorithm is much lower than that of a stress based algorithm.

Another important observation is that the harmony effect in word spotting only emerged when reaction times were very slow. When there was no stress cue, the average RT was 807 ms measured from word offset. This is extremely slow if one considers that, for example, close shadowers often initiate their response before the end of the word is heard (Marslen-Wilson, 1973). It also contrasts with the fact that a stress cue speeded responses by more than 500 ms. A similarly big RT difference was found, but not commented on, by Suomi et al. (1997). Their average word spotting RTs were 731 ms in Experiment 1, but when words were spliced from their context, RTs dropped by 360 ms to an average of 371 ms. The question is how to account for those large overall differences.

One answer may come from the comments of participants performing the word spotting experiments. When there was no stress cue, participants complained that the task was extremely difficult. For many, it was more like a metalinguistic task in which explicit instructions about the nature of the task and the items was required. If participants had not been told that pseudowords contained other embedded words, they would probably not have discovered it at all. This contrasts with the case in which there was a stress cue: The task was very easy, words just ‘‘popped out’’ of the speech signal, and the identity of the embedded word was immediately obvious. These observations strongly suggest that the nature of the task was very different in Experiments 1 and 2. An often made distinction in this respect is the on-line versus off-line nature of a tasks. Word spotting is usually classified as an on-line task, because RTs are measured from participants who are required to make a speeded response. However, it can be questioned whether the speed requirement as such is sufficient, because there are serious reasons to doubt the on-line nature of a task when RTs are extremely slow. We therefore refrain from an unqualified classification of word spotting as an on-line task.

In contrast, the learning task of Experiment 3 is probably considered an off-line task, because speed as such is not a requirement. However, despite its alleged off-line nature, the comparison between language groups allows us to conclude that a language-specific component is tapped that should be highly relevant in on-line speech segmentation. Listeners relied on the rhythmic and phonological characteristics of their native language when segmenting unfamiliar speech input. Thus, Finns profited from vowel harmony and word-initial stress in the same interdependent way as was found in word spotting, Dutch profited from word-initial stress, and French profited neither from vowel harmony nor from word-initial stress. These are, of course, exactly the properties to which one would expect a language-specific segmentation routine to be tuned. It therefore seems that an off-line task can be informative about on-line processing.

Another issue that requires some discussion concerns the role of stress in lexical access. From the present results it is clear that a stressed syllable can signal a word boundary, but this by itself does not imply that stress is part of the lexical input representation. In fact, we prefer to view the status of a stress cue as akin to that of any other phonetic cue that signals a word boundary. The prime example is a long silence: Any speech sound after a silence of, say, 1 s is likely to be the onset of a new word, but this does not imply that the silence itself is part of the lexical representation of the word. In fact, it is very likely that it is not. Silence is thus a reliable segmentation cue, but it is not part of the lexical representation. Similarly, we would argue that a stressed syllable is a reliable segmentation cue for Finnish listeners, but the input representation of the word itself does not distinguish between stressed and unstressed syllables. The reason is simply that stress is not distinctive. In fact, coding stress in the input representation of the word would be completely redundant because each word has stress on its first syllable. From this viewpoint, then, it seems likely that stress is not part of the input representation. This probably allows an unstressed or even mis-stressed word to be recognized as a (mis-stressed) *word*, and not as a *nonword*. Similarly, it may explain why FORbear primes the associate of forBEAR (Cutler, 1986). Word stress is thus not used in the way segmental structure is: It cues a word boundary, but it does not constrain the number of lexical candidates.

In conclusion, the present study showed that Finnish word boundaries are signaled by vowel disharmony and word stress. We argued that stress dominates vowel disharmony because the former is more informative than the latter. It may also be that, during on-line word recognition, stress is available much earlier than vowel disharmony. For example, stressed syllables are more salient, and saliency itself may be perceived quickly. In contrast, vowel disharmony relies on the relation between an unstressed word-final vowel and a stressed word-initial vowel. This is a syntagmatic relation that may be difficult to compute. Word boundary cues may therefore have different time courses at which they become available. This implies that if one wants to obtain a realistic view of how listeners deal with multiple segmentation cues, one needs to study them not only in isolation but also in conjunction.

APPENDIX

Experimental Items and Prefixes Used in Experiments 1 and 2

	Harmonious prefix		Disharmonious prefix		
Harmony class	Prefix	Word	Prefix	Word	Gloss
Back	ku	palo	ky	palo	fire
	ka	kuja	kä	kuja	alley
	po	lato	pö	lato	barn
	tu	haka	ty	haka	hook
	to	luku	tö	luku	number
	pu	juna	py	juna	train
	po	sopu	pö	sopu	agreement
	ku	romu	ky	romu	trash
	po	kuva	pö	kuva	picture
	po	muna	pö	muna	egg
	to	latu	tö	latu	track
	ta	raju	tä	raju	rash
	pu	tupa	py	tupa	cottage
	ku	kora	ky	koru	jewelery
	tu	napa	ty	napa	navel
Front	ty	kynä	tu	kynä	pen
	py	näkö	pu	näkö	sight
	kä	pöly	ka	pöly	dust
	ky	sävy	ku	sävy	shade
	ty	hätä	tu	hätä	emergency
	ky	pyry	ku	pyry	snowfall
	ty	kyky	tu	kyky	ability
	pö	käry	po	käry	odour
	tö	häkä	to	häkä	carbon monoxide
	py	hymy	pu	hymy	smile
	pö	läjä	po	läjä	heap
	tö	käpy	to	käpy	pine cone
	ky	rysä	ku	rysä	trap
	pö	syvä	po	syvä	deep
	tä	tyly	ta	tyly	harsh

REFERENCES

Cole, R., & Jakimik, J. (1980). A model of speech perception. In R. Cole (Ed.), *Perception and production of fluent speech.* Hillsdale, NJ: Erlbaum.

Cutler, A. (1986). *Forbear* is a homophone: lexical prosody does not constrain lexical access. *Language and Speech,* **29,** 201–220.

Cutler, A., Dahan, D., & Donselaar, V. van (1997). Prosody in the comprehension of spoken language: A literature review. *Language and Speech,* **40,** 141–201.

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

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

Dell, F., & Vergnaud, J. -R. (1984). Les développements récents en phonologie: Quelques idées centrales. In F. Dell, D. Hirst, & J. -R. Vergnaud (Eds.), *Forme sonore du langage* (pp 1–42) Paris: Hermann.

Fear, B. D., Cutler, A., & Butterfield, S. (1995). The strong/weak syllable distinction in English. *Journal of the Acoustical Society of America,* **97,** 1893–1904.

Iivonen, A., Niemi, T., Paananen, M. *Do F0 peaks coincide with lexical stresses? A comparison of English, Finnish, and German.* Manuscript submitted for publication.

Kager, R. J. W. (1989). *A metrical theory of stress and destressing in English and Dutch.* Foris, Dordrecht.

Karlsson, F. (1983). *Suomen kielen äänne—ja muotorakenne (Finnish phonological and morphological structure.* Helsinki, WSOY.

Lehiste, I. (1960). An acoustic-phonetic study of internal open juncture. *Journal of the Acoustical Society of America,* **51,** 2018–2024.

Marslen-Wilson, W. D. (1984). Function and process in spoken word recognition. In H. Bouma & D. Bouwhuis (Eds.), *Attention and performance X* (pp. 125–150). Hillsdale, NJ: Erlbaum.

Marslen-Wilson, W. D. (1973). Linguistic structure and speech shadowing at very short latencies. *Nature,* **244,** 522–523.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology,* **18,** 1–86.

McQueen, J. M., Norris, D. G., & Cutler, A. (1994). Competition in spoken word recognition: Spotting words in other words. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* **20,** 621–638.

Nakatani, L. H., & Dukes, K. D. (1977). Locus of segmental cues for word juncture. *Journal of the Acoustical Society,* **62,** 714–719.

Nakatani, L. H., & Schaffer, J. A. (1978). Hearing 'words' without words: Prosodic cues for word perception. *Journal of the Acoustic Society of America,* **63,** 234–245.

Norris, D. G. (1994). SHORTLIST: A connectionist model of continuous speech recognition. *Cognition,* **52,** 189–234.

Norris, D. G., McQueen, J. M., & Cutler, A. (1995). Competition and segmentation in spoken word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* **21,** 1209–1228.

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.

Suomi, K., McQueen, J. M., & Cutler, A. (1997). Vowel harmony and speech segmentation in Finnish. *Journal of Memory and Language,* **36,** 422–444.

't. Hart, J., Collier, R., & Cohen, A. (1990). *A perceptual study of intonation: An experimental approach to speech melody.* Cambridge: Cambridge University Press.

Umeda, N. (1975). Vowel duration in American English. *Journal of the Acoustical Society of America,* **58,** 434–445.

Vroomen, J., & de Gelder, B. (1995). Metrical segmentation and lexical inhibition in spoken word recognition. *Journal of Experimental Psychology: Human Perception and Performance,* **21,** 98–108.

Vroomen, J., & de Gelder, B. (1997). The activation of embedded words in spoken word recognition. *Journal of Experimental Psychology: Human Perception and Performance,* **23,** 710–720.

Vroomen, J., van Zon, M., & de Gelder, B. (1996). Cues to speech segmentation: Evidence from juncture misperceptions and word spotting. *Memory and Cognition,* **24,** 744–755.

(Received March 12, 1997)