JOURNAL OF MEMORY AND LANGUAGE **36,** 422–444 (1997)
ARTICLE NO. ML962495

Vowel Harmony and Speech Segmentation in Finnish

KARI SUOMI

University of Oulu, Oulu, Finland

AND

JAMES M. MCQUEEN AND ANNE CUTLER

Max-Planck-Institute for Psycholinguistics, Nijmegen, The Netherlands

Finnish vowel harmony rules require that if the vowel in the first syllable of a word belongs to one of two vowel sets, then all subsequent vowels in that word must belong either to the same set or to a neutral set. A harmony mismatch between two syllables containing vowels from the opposing sets thus signals a likely word boundary. We report five experiments showing that Finnish listeners can exploit this information in an on-line speech segmentation task. Listeners found it easier to detect words like *hymy* at the end of the nonsense string *puhymy* (where there is a harmony mismatch between the first two syllables) than in the string *pyhymy* (where there is no mismatch). There was no such effect, however, when the target words appeared at the beginning of the nonsense string (e.g., *hymypu* vs *hymypy*). Stronger harmony effects were found for targets containing front harmony vowels (e.g., *hymy*) than for targets containing back harmony vowels (e.g., *palo* in *kypalo* and *kupalo*). The same pattern of results appeared whether target position within the string was predictable or unpredictable. Harmony mismatch thus appears to provide a useful segmentation cue for the detection of word onsets in Finnish speech.

In Finnish, the word for ‘‘sight’’ is *näkö.* The word for ‘‘taste’’ is *maku. Muka*, as it happens, is also a word. But there is no word *näku* in Finnish, and no words *nukä*, *makö,* or *möka*, nor is any of these a possible native Finnish word. This is because there are strict constraints on the occurrence of vowels within a Finnish word: *a* and *u* belong to one class and *ä* and *ö* to another class, and these two classes may not co-occur within a word.

This system is known as vowel harmony. In Finnish, there are two opposing harmony classes: three front vowels, /y, ø, æ/, and three back vowels, /u, o, a/. If a native Finnish word contains a front vowel in its first syllable, all other vowels must also be front (unless they are a member of a so-called neutral set, /i, e/). Similarly, if the first syllable contains a back vowel, all other vowels must be neutral or back.

As observed more than a half century ago by Trubetzkoy (1939), vowel harmony has the potential to serve an extremely important function in speech perception: It could function as a source of information that listeners could exploit to locate the onsets of individual words in the continuous speech stream. Since vowels from opposing harmony sets cannot occur within a word, the presence of a front vowel in a syllable in continuous speech following a syllable with a back vowel signals a word boundary between those two vowels. In this paper, we ask whether Finnish listeners can use vowel disharmonies to assist in speech segmentation.

Why listeners might need to make use of information of this kind is that spoken lan-

The first author thanks Esa Itkonen for his unfailing support and encouragement over the years and the Max Planck Society for a visiting fellowship. We also thank Ilse Lehiste for her comments on an earlier version of this paper and Pentti Körkkö for many kinds of technical help. Address correspondence and reprint requests to Kari Suomi, Logopedics, Phonetics and Speech Communication, University of Oulu, P.O.B. 111, 90571 Oulu, Finland. E-mail: kari.suomi@oulu.fi.

0749-596X/97 $25.00

guage lacks reliable and deterministic cues to word boundaries (Lehiste, 1972; Nakatani & Dukes, 1977). In contrast with written language, where alphabetic scripts in general indicate boundaries with white spaces between words, speech provides no such mandatory signals. Nevertheless, listeners apparently effortlessly segment and recognize the words embedded in continuous utterances. Vowel harmony may provide speakers of languages such as Finnish with one means by which they achieve this effortless performance.

A solution to the segmentation problem is, in principle, offered by models of word recognition that are based on active competition between multiple lexical hypotheses (Shortlist: Norris, 1990, 1994; TRACE: McClelland & Elman, 1986). In Shortlist, it is assumed that words can begin at any point in the speech input. Only those words that best match the currently available input are considered (a ‘‘short list’’ of candidate words). Word hypotheses that are partially activated by the same input segments compete with one another through a process of lateral inhibition, until the model settles on an optimal interpretation of the input.

This competition process allows the model to select words (and hence segmentations) from a continuous input. Consider the British English phrase ‘‘ship inquiry,’’ with no explicit marking of its internal word boundary. Candidate words in Shortlist for this phrase would include *ship, shipping, inquiry,* and *choir.* In the model, *ship* would compete with *shipping,* and *inquiry* with both *shipping* and *choir.* The candidate *inquiry* (in British English, stressed on the second syllable) will inhibit *choir* because it receives more support from the input (i.e., it accounts for the final /ri/), and as the activation of *inquiry* increases, it will compete more effectively with *shipping.* As the activation of the *shipping* candidate is thus reduced, the activation of *ship* (also competing with *shipping*) will rise. The model therefore settles on the *ship inquiry* reading, that is, on the correct segmentation of the input. Segmentation of continuous speech thus emerges as a consequence of the competition process (for more extended discussion, see McQueen, Cutler, Norris, & Briscoe, 1995; Norris, 1994; Norris, McQueen, & Cutler, 1995). Recognition and segmentation through competition is supported by a growing body of evidence, from several languages: English (Cluff & Luce, 1990; Goldinger, Luce, Pisoni, & Marcario, 1992; McQueen, Norris, & Cutler, 1994; Norris et al., 1995); Italian (Tabossi, Burani, & Scott, 1995); and Dutch (Vroomen & de Gelder, 1995).

Despite the potential efficiency of the competition process in segmentation, there is clear empirical evidence that listeners also make use of information in the signal. One such type of information, for instance, is the metrical structure of speech. Evidence from many languages attests to listeners’ use of metrical information. In stress-timed languages like English and Dutch, the rhythmic distinction between strong and weak syllables appears to be used by listeners in segmentation (Cutler & Butterfield, 1992; Cutler & Norris, 1988; McQueen et al., 1994; Norris et al., 1995; Vroomen & de Gelder, 1995; Vroomen, van Zon, & de Gelder, 1996). In syllable-timed languages like French, Catalan, and Spanish, listeners appear to use the syllable for segmentation (Cutler, Mehler, Norris, & Seguí, 1986, 1992; Mehler, Dommergues, Frauenfelder, & Seguí, 1981; Pallier, Sebastián-Gallés, Felguera, Christophe, & Mehler, 1993; Sebastián-Gallés, Dupoux, Seguí, & Mehler, 1992). In Japanese, the rhythmic unit is the mora, and Japanese listeners appear to use morae for segmentation (Cutler & Otake, 1994; Otake, Hatano, Cutler, & Mehler, 1993).

Note that because languages differ rhythmically, different metrical information must be used depending on the input language. Whatever the rhythmic unit, the basic claim remains the same: Word onsets will tend to be aligned with the onsets of metrical units, so if a boundary between two metrical units can be detected in the signal, it is likely that it will correspond to a word boundary. The detection of likely word-onset locations using metrical information can thus aid the recognition process. The use of such information is fully compatible with an ongoing competition process in word recognition; indeed, the experiments by McQueen et al. (1994) and Norris et al. (1995) provided concomitant evidence both for competition and for metrically based segmentation.

In addition to the cues provided by rhythm, languages may differ with respect to other resources that can be used to assist in segmentation. Although phonetic cues to word boundaries such as lengthening of onset syllables and segments (Gow and Gordon, 1995; Lehiste, 1972) and aspiration of word-initial stops (in English, Nakatani & Dukes, 1977) are indeterminate (they do not occur reliably), this does not mean that they could not be used to aid segmentation when they are available. Church (1987) has proposed that several different types of phonological knowledge, including the phonotactic and allophonic cues to syllable structure (varying across languages), could be used to improve the efficiency of the recognition process. Phonotactic information is of use as a segmentation cue, as has recently been demonstrated in studies using the word-spotting task (McQueen & Cox, 1995). In word spotting (Cutler & Norris, 1988), the listener's task is to detect words embedded in nonsense strings. McQueen and Cox (1995) showed that words were easier to spot in bisyllabic nonsense strings when they were aligned with syllable boundaries than when they were misaligned with syllable boundaries. Crucially, these syllable boundaries were determined by phonotactic constraints. In /fi.drɒk/, for example, there must be a syllable boundary before the /d/ (because voiced stops do not occur in coda position in Dutch). In /fim.rɒk/, however, the boundary must occur between the /m/ and the /r/ (an /mr/ cluster is illegal in Dutch). Dutch listeners were slower and less accurate in detecting *rok* (skirt) in /fi.drɒk/ than in /fim.rɒk/. Allophonic variation is also of use as a segmentation cue. It can be used to distinguish minimal pairs in French (Zwanenburg, Ouweneel, & Levelt, 1977), in Japanese (Shimizu & Dantsuji, 1980), in English (Christie, 1974; Nakatani & Dukes, 1977), and in Dutch (Quené, 1987). Furthermore, Yerkey and Sawusch (1993), again using the word-spotting task, have shown that allophonic cues such as the aspiration of stops can influence listeners' ability to detect words embedded in nonsense strings.

The recognition and segmentation of continuous speech can therefore be seen as a process based on lexical competition, but one that uses whatever additional segmentation cues the language in question provides, whenever those cues are available in the speech signal. Moreover, cues may be exploited even when their information value is only partial—for example, when they are capable of signaling only the presence but not the absence of a boundary.

Vowel harmony in Finnish is just such a potential segmentation cue. One of the more common types of vowel harmony in the world's languages is palatal (front-back) harmony, which occurs most extensively in Uralic and Altaic languages, e.g., Finnish and Turkish (Turkish also has labial harmony). In these languages, harmony propagates left to right from the first vowel in the root to subsequent vowels in the root and in suffixes, but not across word boundaries. Of the Finnish vowel phonemes /i e y ø æ a o u/ (unambiguously represented in the orthography by 〈i e y ö ä a o u〉, respectively), the set /u o a/ are Back Harmonic, the set /y ø æ/ are Front Harmonic, and /i e/ are Harmonically Neutral. Note that the vowels in the last two classes are all phonetically front and that the vowels in the two harmonic classes are pairwise distinguished by the backness feature alone (/u - y/, /o - ø/, /a - æ/). Finnish also has a vowel quantity distinction, such that all eight vowels can occur either singly or doubled.

The main restriction imposed by Finnish vowel harmony is that, within an uncompounded word form, vowels from only one of the two harmonic classes can occur, whereas the harmonically neutral vowels may be combined with vowels from either harmonic class in any word position. If a word contains derivational and/or inflectional suffixes, any vowels in these suffixes are also subject to the harmony restriction. Thus, there are words like *osuma, pesula, kupari, seteli,* and *kypärä, veräjä, kätevä, rypäle,* whereas words of the type **osyma, *pesyla, *kypära* do not occur in the native vocabulary. In the past, borrowed words violating vowel harmony were always adapted to the native pattern, although nowadays many borrowed words exist that are not adapted (e.g., *parfyymi, dynastia, volyymi*). Despite the existence of such exceptions, the occurrence in speech of disharmonious vowels in consecutive syllables suggests that those vowels belong to two separate words, and hence that there is a word boundary between the vowels.

In Experiment 1, we attempted to ascertain whether vowel harmony could indeed, as suggested by numerous linguists (Karlsson, 1983; Trubetzkoy, 1939), be exploited by listeners in segmentation. The word-spotting task was used. Listeners were presented with trisyllabic nonsense strings, some of which ended with embedded disyllabic words. For example, the target *hymy* was embedded in the strings *pyhymy* and *puhymy.* The first of these strings is harmonious in that all vowels belong to the same (front) harmonic class. In the second string, in contrast, the vowel in the initial syllable is disharmonious with those in the embedded word. Back Harmonic words were similarly embedded in harmonious and disharmonious contexts (e.g., *palo* in *kupalo* and *kypalo*). If speakers of Finnish can exploit vowel harmony in speech segmentation, they should find it easier to spot target words in disharmonious than in harmonious strings.

EXPERIMENT 1

Method

Materials. Thirty CVCV words were chosen, half of them containing only back harmonic vowels, the other half only front harmonic vowels. The words are all monomorphemic nouns or adjectives in their singular nominative (i.e., basic, uninflected) form. Examples are *palo* (fire) and *hymy* (smile). None of the chosen words contained other words embedded in them. Two alternative CV contexts were prefixed to each word to create trisyllabic nonwords. None of these CVs were words. Although single CVs are rare as words in Finnish, they are possible words: a small set of function words, including the conjunction *ja* (and) and the pronoun *me* (we) do exist. For each word, one CV context had a vowel belonging to the same harmonic class as the vowels in the embedded word, the other had a vowel from the opposite harmonic class. Thus, each embedded word occurred in the context of a harmonic initial vowel, and in the context of a disharmonic initial vowel. For example, to *palo* were added both *ku* and *ky,* and to *hymy* both *py* and *pu.* This pairwise addition of CV contexts produced 60 trisyllabic items, none of which contained, besides the intended target word, any other words, nor could the items be continued to form longer noncompound words. Because it was not possible to obtain reliable frequency of occurrence information for the target words, they were selected by the first author on the basis of his own familiarity judgments. No highly unfamiliar words were used, and the back and front harmony words were matched on familiarity. The first author also checked the words for their uniqueness points. By his judgment, 23 of the words became unique on their final phonemes, and the remaining 7 did not become unique until after their offsets.

Because of the very stringent constraints on the choice of the target words, and especially because of the pairwise addition of context syllables to each target word, it was not possible to control the number of words compatible with the initial CV sequence separately for each carrier pair (e.g., the number of words compatible with the contexts *ku* and *ky* in the pair *kupalo, kypalo*). In all of the disharmonic materials and in 47% of the harmonic materials, the first two syllables could not be continued to form any words. In addition, across the total sets of harmonic and disharmonic contexts, an exactly equal number of each individual CV syllable occurred. Thus, for example, *py* occurred twice both as a harmonic context and as a disharmonic context. Vowel height was kept constant across the members of each pair of first vowels; that is, /u/ was always paired with /y/, /o/ with /ø/, and /a/ with /æ/. In both the target words and context syllables (and in the fillers) all vowels were single. The target-bearing items are listed in the Appendix.

Eighty trisyllabic CVCVCV nonword fillers were also constructed. In half of these filler items the vowels in the last two syllables were Back Harmonic, in the other half they were Front Harmonic. Within both sets, half of the items had a first vowel that was harmonious with the vowels in the last two syllables, while in the other half the first vowel was disharmonious with the following vowels. None of the filler items contained a word, including inflected word forms, nor could they be continued to form longer words.

The materials were recorded by a male speaker of Finnish, a phonetician ignorant of the purpose of the recording, using a DAT recorder. A single recording of the whole set of materials was made. All items were produced with the typical prosody of Finnish trisyllabic CVCVCV words. Two lists were then constructed, both containing all of the 80 nonword filler items, and all of the target words, each of which appeared in a given list in either a harmonic or a disharmonic context; type of context was counterbalanced over lists. In their respective lists, the members of each harmonic–disharmonic pair occupied the same serial position. A set of 5 practice items was placed at the beginning of both lists.

Subjects. Twenty-eight voluntary subjects, logopedics or language students at Oulu University, Oulu, Finland, took part. Fourteen subjects heard each list.

Procedure. Subjects were tested individually. They were seated in front of a computer in a quiet room, and the materials were presented through headphones. The interval between consecutive items was 3 s. Subjects were instructed that they would hear nonsense items that could contain finally embedded real words, and that they should press the response key whenever they heard a nonsense word ending in a real word. They were then to say aloud the word they had detected. These spoken responses were recorded on a video cassette recorder.

The experimental items were stored as separate files at 20-kHz sampling rate and 10-bit resolution on the hard disk of a Compaq Deskpro 386/20e computer. They were played to subjects directly from disk, under the control of EASYST, a DOS-compatible reaction time measurement system constructed by Einar Meister at the Phonetics and Speech Technology Laboratory of the Estonian Academy of Sciences in Tallinn (Meister & Suomi, 1993). Reaction times were measured by an external timer with a tested resolution of less than 1 ms and stored on the computer. Target durations were measured using EASYST.

Results and Discussion

Reaction Times (RTs), originally measured from the onset of each trisyllabic item, were adjusted by subtracting total item durations, yielding RTs from target word offset. Subjects' spoken responses were then analyzed, and it was found that whenever subjects made a response to a target-bearing item, they detected the intended target word. That is, all button-press responses were associated with correct oral responses. Outlying responses (those faster than 150 ms or slower than 2000 ms, as measured from target offset) were treated as errors (7% of the data). In most word-spotting studies, subjects have been excluded from the analysis if they failed to reach a criterial level of performance and items have been excluded if they were missed by too many subjects. In this and all subsequent experiments we adopted the criteria that each subject should detect at least 50% of the targets they heard and that each item should be detected by at least 50% of the subjects who heard it. In Experiment 1, no subjects or items were rejected on the basis of these criteria. The mean RTs and error rates are shown in Table 1.

Analyses of variance (ANOVAs) were performed on RTs and errors with both subjects ($F1$) and items ($F2$) as the repeated measure. Words in disharmonious strings (e.g., *hymy* in *puhymy*) were detected, on average, 161 ms

TABLE 1

MEAN REACTION TIME (RT, IN MS) MEASURED FROM TARGET WORD OFFSET AND MEAN PERCENTAGE ERROR RATES (ERR) IN EXPERIMENT 1

	Target harmony class					
	Back		Front		Overall	
Context	RT	Err	RT	Err	RT	Err
Harmonious	802	9%	822	10%	812	10%
Disharmonious	699	5%	604	4%	651	5%

faster than words in harmonious strings (e.g., *hymy* in *pyhymy*): $F1(1,26) = 48.61$, $p < .001$; $F2(1,28) = 28.40$, $p < .001$. This effect interacted with the harmony class of the target word: $F1(1,26) = 15.83$, $p < .001$; $F2(1,28) = 5.87$, $p < .05$. The effect was larger for words with front harmony vowels than for words with back harmony vowels. In comparisons within harmony class, the advantage for front harmony words in a disharmonious context (e.g., *hymy* in *puhymy*) over those in a harmonious context (e.g., *hymy* in *pyhymy*) was reliable: $t1(27) = 10.50$, $p < .001$; $t2(14) = 5.75$, $p < .001$ (218 ms on average). The advantage for back harmony words in a disharmonious context (e.g., *palo* in *kypalo*) over those in a harmonious context (e.g., *palo* in *kupalo*) was reliable by subjects but marginal by items: $t1(27) = 2.92$, $p < .01$; $t2(14) = 1.97$, $p = .07$ (103 ms, on average). Overall, responses to the front harmony targets were faster (38 ms, on average) than those to the back harmony targets, but this was only significant by subjects: $F1(1,26) = 5.30$, $p < .05$; $F2 < 1$.

In the analysis of errors, there was also an advantage for targets in disharmonious strings, which were detected, on average, 6% more accurately than targets in harmonious strings: $F1(1,26) = 10.72$, $p < .005$; $F2(1,28) = 5.18$, $p < .05$. No other main effects or interactions were significant in this analysis.

The target durations were submitted to an ANOVA. Target words in harmonious strings were found to be 19 ms shorter (374 ms, on average) than those in disharmonious strings (393 ms, on average): $F(28) = 10.81$, $p < .005$. This confound, however, cannot account for the harmony effect. In correlational analyses comparing mean RT and word duration over items, there was no reliable evidence to suggest that responses were faster to longer words. There were no significant correlations of RT with duration for either front or back harmony targets in harmonious strings nor for back harmony targets in disharmonious strings, but there was a negative correlation (faster responses to longer words) for front harmony targets in disharmonious strings: $r(14) = -.54$, $p < .05$. There were no reliable correlations of target duration and error rate.

As a further control for effects of target duration, the data were reanalysed, measuring RT from word onset (now counting as outliers those outside the range of 150–2000 ms as measured from word onset; 9% of the data). Responses to targets in disharmonious strings were 129 ms faster, on average (mean 1027 ms), than responses to targets in harmonious strings (mean 1156 ms): $F1(1,26) = 41.73$, $p < .001$; $F2(1,28) = 26.49$, $p < .001$. As in the offset analysis, this effect interacted with the harmony class of the target word, but the interaction was only reliable by subjects: $F1(1,26) = 8.81$, $p < .01$; $F2(1,28) = 3.54$, $p = .07$. Pairwise comparisons confirmed that the harmony effect was reliable in this analysis both for front-harmony words (e.g., *hymy* in *puhymy* and *pyhymy*, mean difference 166 ms): $t1(27) = 8.63$, $p < .001$; $t2(14) = 5.43$, $p < .001$; and for back-harmony words (e.g., *palo* in *kypalo* and *kupalo*, mean difference 96 ms): $t1(26) = 2.97$, $p < .01$; $t2(14) = 2.14$, $p = .05$. With these data, there were no reliable correlations of target duration with RT. Target duration does not appear to be responsible for the overall differences in word-spotting performance between harmonious and disharmonious strings.

A final analysis explored effects of lexical competition. In all of the disharmonious contexts, there were no words consistent with the first two syllables of the nonsense strings. In almost half of the harmonious contexts, this was also the case. In some strings, however, the first two syllables could be continued to form longer words (e.g., the word *kupari,* copper, overlaps in the first two syllables with *kupalo*). The harmony effect could be due to effects of competition (e.g., *palo* more difficult to detect in *kupalo,* because of competition from *kupari,* than in *kypalo*). An analysis based on RTs from word offsets separated responses to targets that in harmonious contexts had such competitors from responses to other targets. Although the difference between harmonious and disharmonious contexts was larger for targets with competitors in the harmonious context (175 ms, on average) than for targets with no competitors in either context (155 ms, on average), there was no significant interaction, neither by subjects nor items, of this competition factor with the harmony effect. There was also no significant interaction in an error rate analysis (a mean difference of 5% between detection accuracies in harmonious and disharmonious strings in both competitor conditions). Competition effects therefore cannot account for the harmony effect. Competitors, if considered, can be ruled out when they mismatch the final syllables of the nonsense strings (e.g., *kupari* mismatches with the *lo* of *kupalo*) such that they do not reliably influence detection of targets.

These results suggest strongly that disharmonies can be used by Finnish listeners to assist in segmentation. In both RT and errors, listeners showed that they could detect targets in disharmonious strings more easily than targets in harmonious strings. The effects appear to be larger for targets from the front harmony class, such as *hymy,* than for targets from the back harmony class, such as *palo.* Although the RT effect was larger for the front harmony words, and not fully reliable for back harmony words in the analysis from word offset, the back harmony RT effect reached significance in the analysis from word onset. Furthermore, the error rate effect was equivalent across harmony classes.

EXPERIMENT 2

Experiment 1 shows that a harmony mismatch between the first syllable and the following two syllables in a trisyllabic string appears to signal a word boundary. In this situation, the harmony mismatch provides a segmentation cue at the onset of the target word. As Cutler and Norris (1988) argued, cues to the location of word onsets are much more important for word recognition than cues to word offsets. Once a word has been accessed, the location of its offset can be determined by the lexicon (e.g., *palo* must end after the /o/). Experiment 1 thus shows that vowel harmony provides a segmentation cue where it matters most: at the beginning of a word. But could a harmony mismatch also provide a segmentation cue at word offsets?

We addressed this question in Experiment 2. We asked listeners to spot bisyllabic target words at the beginning of trisyllabic nonsense strings, instead of at the end, as in Experiment 1. The same target words were used, but with harmonious and disharmonious context syllables following the targets (e.g., *hymy* in *hymypy* and *hymypu*). If a harmony mismatch at the offset of a word can be used to assist in recognition of that word, word-spotting performance should be easier in disharmonious strings, as in Experiment 1. If, however, segmentation cues primarily signal word onsets, there may be no difference in listeners' ability to spot words in harmonious and disharmonious strings.

Method

Materials. The same 30 target words were used as in Experiment 1 but this time embedded initially in their carrier items. That is, two alternative CV syllables, one harmonic and the other disharmonic, were added to the end of each target word to make two trisyllabic nonsense items (see Appendix). Where possible, the CV contexts used in Experiment 1 were moved from the beginnings of the target words to their ends. Thus, while *palo* was embedded in *kupalo* and *kypalo* in Experiment 1, it was embedded in *paloku* and *paloky* here. However, this simple move was not possible for all target words, because in some cases it would have created other embedded words. For example, *tupa* (cottage) was embedded in *putupa* in Experiment 1, but the item *tupapu* would contain three embedded words: *papu* (bean) and *apu* (help), in addition to the target *tupa*. Moreover, in many cases *t*V endings could not be used, because the combination target + *t* would constitute the plural nominative form of the target (e.g., *kynät*, pens), and endings *pa* and *ko* and their Front Harmonic alternants were avoided altogether because they are enclitics that could be added to any of the target words. While the contexts added to the target words were thus not all identical with those used in Experiment 1, it remained the case that each individual CV syllable occurred equally often in the harmonic set of endings and in the disharmonic set; e.g., *py* occurred twice both as a harmonic and as a disharmonic ending. Vowel height across the members of each pair of contextual harmonic and disharmonic CV syllables was again kept constant.

Note that it was with a view to Experiment 2 that the context syllables started with plosives in both Experiments 1 and 2. Since RTs were to be measured from target word offset, a reliable measuring point was required at this point (also the onset of the following syllable). Plosives were selected since the onset of the release burst is usually easily locatable in acoustic displays.

Eighty fillers, not containing embedded words, were again constructed to be phonologically and phonotactically similar to the target-bearing items. Thus, in half of the fillers the vowels in the first two syllables were Back Harmonic, in the other half they were Front Harmonic, and within both sets, half of the items had a final vowel that was harmonious with the vowels in the first two syllables, the other half a vowel that was disharmonious with the preceding vowels. As in the target-bearing items, the final CV syllable had a plosive onset.

The materials were recorded as in Experiment 1, by the same speaker, and again with the typical prosody of Finnish trisyllabic CVCVCV words. Two lists were again constructed, with exactly the same structure as in Experiment 1.

TABLE 2

MEAN REACTION TIME (RT, IN MS) MEASURED FROM TARGET WORD OFFSET AND MEAN PERCENTAGE ERROR RATES (ERR) IN EXPERIMENT 2

	Target harmony class					
	Back		Front		Overall	
Context	RT	Err	RT	Err	RT	Err
Harmonious	513	9%	492	11%	502	10%
Disharmonious	516	14%	500	7%	508	10%

Subjects. Thirty-eight voluntary subjects, again logopedics or language students at Oulu University, took part; none of them had participated in Experiment 1. Nineteen subjects heard each list.

Procedure. The procedure was the same as in Experiment 1, except that subjects were instructed that they would hear nonsense items that could contain initially embedded real words, and that they should press the response key whenever they heard a nonsense word beginning with a real word. Target durations were again measured using EASYST.

Results and Discussion

RTs were measured from the release burst of the plosive in the onset of the third syllable of each string, as an estimate of the acoustic offset of the target word. As in Experiment 1, it was found in an analysis of the subjects' spoken responses that whenever subjects responded to target-bearing strings, they detected the intended targets. Outlying responses (those faster than 150 ms or slower than 2000 ms, as measured from target offset) were treated as errors (3% of the data). No subjects or items failed the exclusion criteria. The mean RTs and error rates are shown in Table 2.

ANOVAs were again performed on both RTs and error rates, with both subjects and items as the repeated measure. There was no harmony effect: Targets were detected equally quickly and equally accurately in harmonious and disharmonious strings ($F1$ & $F2 < 1$, in both RT and error analyses). Responses to front harmony words were somewhat faster

(19 ms, on average) than those to back harmony words, but this difference was not significant: $F1(1,36) = 2.42$, $p > .1$; $F2 < 1$. No other effects were reliable in the RT analysis. In the error analysis, there was an interaction of harmony context with the harmony class of the target (for targets with back vowels, there were more errors on disharmonious strings, 14% on average, than on harmonious strings, 9% on average; the reverse was true for targets with front vowels, with means of 11 and 7% for harmonious and disharmonious strings, respectively). But this effect was only significant by subjects: $F1(1,36) = 5.81$, $p < .05$; $F2(1,28) = 2.04$, $p > .1$. No other effects were reliable in the error analysis.

There was no difference in measured target durations. Targets in harmonious contexts (mean 326 ms) were of equivalent length to those in disharmonious contexts (mean 320 ms): $F(1,28) = 1.77$, $p > .1$.

In contrast with Experiment 1, where there were strong harmony effects at target word onsets, harmony mismatches at word offsets did not influence performance in Experiment 2. It therefore appears that harmony information only acts as a segmentation cue when it marks a word onset, that is, at a point in time when the word is being accessed. When a harmony mismatch provides a cue about where a word ends (as in Experiment 2), that word has already been accessed, and its end can therefore be determined by lexical information. Harmony information in the signal may therefore not be needed to signal word offsets.

An alternative explanation, however, is that the task demands of Experiment 2 encouraged listeners to ignore the final syllable. They knew in advance that all target words would be in initial position in the nonsense strings. They could thus focus attention on the first two syllables and ignore the subsequent information. If this were the case, there could be no harmony effect. Notice that the responses in Experiment 2 were very fast, about 250 ms faster, on average, than those in Experiment 1 (as observed in other word-spotting studies varying target position; McQueen et al., 1994). Perhaps the Experiment 2 listeners effectively recognized the targets without processing the context syllable, that is, by ignoring the information in the final syllable. This attentional strategy, though possible, is less likely in Experiment 1, where the context syllable was heard before the target words. Because of this temporal sequence, Experiment 1 listeners could hardly avoid processing the context syllable.

If the listeners' task were to be made more difficult, such that they did not know where target words would occur, they would not be able to use an attentional strategy and would be forced to analyze the nonsense strings more fully. This was the approach taken in Experiment 3. The same materials were again used in a word-spotting task, but target location was mixed such that targets could occur either in initial position (as in Experiment 2) or in final position (as in Experiment 1). Listeners were again told to spot words in nonsense strings but were also told that the words could be either at the beginning or the end of the strings. They would thus be unable to focus selectively on either initial or final position. If the difference between the results of Experiments 1 and 2 were due to use of an attentional strategy in Experiment 2, harmony mismatch effects should be found for both initial- and final-position targets. If, however, the difference were due to a genuine asymmetry in the role of harmony information (effects where disharmonies cue word onsets but not word offsets), harmony mismatch effects should be found for the final-position targets (as in Experiment 1) but not for the initial-position targets (as in Experiment 2).

EXPERIMENT 3

Method

Materials. In Experiment 3 each target word occurred in four different embedding conditions, namely, in items with an added harmonic and disharmonic initial CV syllable, as well as in items with an added harmonic and disharmonic final CV syllable. That is, the target words occurred in conditions corresponding to those of Experiment 1 and to those of Experiment 2. Thus, for example, *palo* occurred in the carrier items *kupalo, kypalo, paloku,* and *paloky.* Context syllables were selected such that they were identical in both positions. This meant that some preceding contexts were not those that were used in Experiment 1; instead, they were replaced with those used as following contexts in Experiment 2. The four items with more than one preceding context syllable are listed with their additional syllables in the Appendix.

This set of conditions calls for four counterbalanced lists. To make the number of words in each of the four embedding conditions equal, two further common words were added to the 30 target words used in Experiments 1 and 2, namely, the Back Harmonic *maku* (taste) and the Front Harmonic *jyvä* (grain). With the resulting 32 words, there could be eight examples of each of the four conditions in each of the four lists.

A new recording was made of all the experimental items and of all the filler items used in Experiments 1 and 2. For Experiment 3, 40 of the fillers from Experiment 1 and 40 of the fillers from Experiment 2 were used. The materials were recorded as in Experiment 1. Four lists were constructed, each containing the 80 filler items and each of the 32 target words in one of the four embedding conditions. Position of embedding (initial or final) and type of context syllable (harmonic or disharmonic) were counterbalanced across lists. As in the earlier experiments, a target word always occupied the same serial position within a list.

Subjects and procedure. Thirty-two volunteer subjects, none of whom had participated in the previous experiments, took part in the experiment. Eight subjects heard each list. The procedure was the same as in Experiments 1 and 2, except that subjects were instructed that they would hear nonsense items that could contain either initially embedded or finally embedded real words, and that they should press the response key whenever they heard a nonsense word either beginning or ending with a real word. Target durations were again measured using EASYST.

Results and Discussion

All RTs were originally measured from the string onset. These raw RTs were corrected, by subtracting either target word duration (for the initial targets) or total string duration (for the final targets), so as to effectively measure from target-word offset. In the analysis of the subjects' spoken responses, five responses were found to be incorrect (subjects responded with words other than the intended targets). These five responses were treated as errors. Outlying responses (those faster than 150 ms or slower than 2000 ms, as measured from target offset) were also treated as errors (20% of the data). These cutoff values led to a large number of data points being rejected but were chosen for compatibility with Experiments 1 and 2. A large number of responses were indeed extremely slow, indicating that listeners found this version of the word spotting task considerably more difficult than the versions used in the earlier experiments (word spotting with mixed target location has previously been found to be more difficult, McQueen et al., 1994). Nevertheless, all subjects passed the exclusion criterion by detecting at least 50% of the targets they heard. Seven targets, however, were detected less than 50% of the time in at least one of the four conditions: *romu, latu, tupa, raju,* and *maku* from the back-harmony set and *rysä* and *tyly* from the front-harmony set. These items were removed from the analysis, leaving 11 back-harmony and 14 front-harmony targets. The mean RTs and error rates are shown in Table 3.

ANOVAs were again performed on both RTs and error rates, with both subjects and items as repeated measures. In the RT analysis, there was a main effect of harmony. Targets were detected, on average, 93 ms faster in disharmonious strings than in harmonious strings: $F1(1,28) = 16.38$, $p < .001$; $F2(1,23) = 8.09$, $p < .01$. But this effect interacted with target position: The effect was large (207 ms, on average) for targets in final position (even larger than in Experiment 1) but in the wrong direction for targets in initial position (a 21 ms-advantage, on average, for the har-

TABLE 3

MEAN REACTION TIME (RT, IN MS) MEASURED FROM TARGET WORD OFFSET AND MEAN PERCENTAGE ERROR RATES (ERR) IN EXPERIMENT 3

	Target harmony class					
	Back		Front		Overall	
	Initial targets					
Context	RT	Err	RT	Err	RT	Err
Harmonious	707	11%	685	6%	696	8%
Disharmonious	721	9%	714	11%	717	10%
	Final targets					
Context	RT	Err	RT	Err	RT	Err
Harmonious	1024	26%	1101	22%	1062	24%
Disharmonious	906	22%	805	12%	855	15%

monious contexts): $F1(1,28) = 18.48$, $p < .001$; $F2(1,23) = 14.65$, $p < .001$. Accordingly, separate ANOVAs were carried out for responses to initial and final targets.

In the final-position analysis, there was a reliable harmony effect, with much faster responses to targets in disharmonious strings than to targets in harmonious strings: $F1(1,28) = 21.09$, $p < .001$; $F2(1,23) = 14.15$, $p < .005$. But this effect interacted strongly with target harmony class: $F1(1,28) = 24.80$, $p < .001$; $F2(1,23) = 5.01$, $p < .05$. The effect was very large indeed for targets with front vowels (e.g., *hymy* in *puhymy* and *pyhymy*, a mean difference of 296 ms): $t1(31) = 5.56$, $p < .001$; $t2(13) = 4.13$, $p < .005$. The effect was smaller, and only significant by subjects, for targets with back vowels (e.g., *palo* in *kypalo* and *kupalo,* a mean difference of 118 ms): $t1(31) = 2.02$, $p = .05$; $t2(10) = 1.18$, $p > .1$. No other effects were significant. In the initial-position analysis, there were no reliable effects.[1]

Similar patterns were observed in the error rates. In the overall analysis, there was no reliable harmony effect, but there was an interaction of harmony with target position, significant by subjects: $F1(1,28) = 4.80$, $p < .05$; $F2(1,23) = 2.46$, $p > .1$. The error rates for each position were analysed separately. In the final-position analysis, the small advantage for targets in disharmonious strings (9%, on average) was significant by subjects, but not by items: $F1(1,28) = 5.78$, $p < .05$; $F2(1,23) = 3.09$, $p = .09$. No other effects were significant. As in the RT analysis, there were no reliable effects in the initial-position error rate analysis.

Measured target durations were also submitted to an ANOVA (excluding the seven items rejected in the above analyses). Targets in disharmonious strings were reliably longer than those in harmonious strings (mean durations: 384 ms, disharmonious; 366 ms, harmonious): $F(1,23) = 11.52$, $p < .005$. This differ-

---

[1] Because a large amount of data was excluded (20% as outliers, then all responses to seven items), a control RT analysis was performed to check that the pattern of results held over all the data. The same harmony mismatch effect was observed in an analysis in which no outliers and no items were excluded ($F1(1,28) = 19.25$, $p < .001$; $F2(1,30) = 11.0$, $p < .005$). This effect interacted with position of the target word ($F1(1,28) = 26.89$, $p < .001$; $F2(1,30) = 14.17$, $p < .001$). As in the main analysis, the harmony effect was limited to the targets in final position (mean RTs from target offset: initial position, 754 ms harmonious, 772 ms disharmonious; final position, 1277 ms harmonious, 1004 ms disharmonious). Similar analyses on the data from Experiments 1 and 4 with no outliers or items excluded also confirmed the reliability of the harmony mismatch effect (which was significant by both subjects and items in both cases).

ence was due almost entirely to front harmony targets (mean durations: 389 ms, disharmonious; 361 ms, harmonious), and not to back harmony targets (mean durations: 378 ms, disharmonious; 372 ms, harmonious), as shown by a significant interaction of harmony and harmony class: $F(1,23) = 4.68$, $p < .05$. Correlations of mean RT and target duration, and of mean error rate and target duration were therefore carried out for the final-position targets. No correlations were significant. No reliable differences in target duration were found for the initial-position targets, and there were no significant correlations of initial-position target durations with either RT or error rate.

Finally, as in Experiment 1, the responses to final targets were examined for effects of lexical competition. Responses to targets such as *palo* which in harmonious contexts had competitors like *kupari* which overlapped with the first two syllables of the string (*kupalo*) were separated from responses to other targets. Although in the RT analysis there was a significant interaction of this competition factor with the harmony effect by subjects ($F1(1,28) = 6.97$, $p < .05$), the interaction was not significant by items ($F2(1,21) = 2.01$, $p > .1$). Note, however, that this interaction was the reverse of that predicted by competition: The difference between harmonious and disharmonious contexts was smaller for targets with competitors in the harmonious context (102 ms, on average) than for targets with no competitors in either context (254 ms, on average). The interaction was not significant by subjects or items in the error analysis (a mean difference of 9% for targets with competitors in the harmonious context and of 7% for targets without competitors). As in Experiment 1, therefore, the harmony effect cannot be due to this competitor asymmetry.

The results of Experiment 3, with a mixed design of both initial and final targets, confirm the pattern observed across Experiments 1 and 2. There were strong harmony mismatch effects for final targets, where the mismatch marks the onset of the target word, in both Experiments 1 and 3. This effect thus emerges both when listeners can focus attention on a specific target location (Experiment 1) and when they cannot (Experiment 3). On the other hand, there were no such harmony mismatch effects for initial targets, where the mismatch marks the offset of the target word (Experiments 2 and 3). It is unlikely that the failure to find an effect in Experiment 2 was due to subjects' ignoring the contextual information in the final syllable, since in Experiment 3 they were required to process the information in all three syllables. It appears, instead, that there is a genuine asymmetry in the harmony effect. Disharmonious information cues word boundaries, but this segmentation cue only assists in the recognition of words following such a boundary, not in the recognition of words preceding such a boundary.

Another feature of the results that was replicated in Experiment 3 was the interaction in the size of the harmony mismatch effect between targets with front-harmony vowels (e.g., *hymy*) and targets with back-harmony vowels (e.g., *palo*), first observed in Experiment 1. In Experiment 1, the mismatch effect was present in the error rates for both back- and front-harmony targets but was reliably larger in the RTs for front- than for back-harmony targets. This interaction was stronger for the final-position targets in Experiment 3. There were reliable effects for front-harmony targets (e.g., *hymy* in *puhymy* and *pyhymy*) in both speed and accuracy, but only a marginal effect for back-harmony targets (e.g., *palo* in *kypalo* and *kupalo*) in RT and no effect in errors. Perhaps this weaker effect was due to the recording used in Experiment 3. On the other hand, the weaker effect for back targets found in Experiment 3 may be due to the mixed-position design. Experiment 4 explored this back/front interaction further. Experiment 4 was a straight replication of Experiment 1, that is, with targets only appearing in final position. The same target words were tested as in Experiment 1, but the recording made for Experiment 3 was used.

EXPERIMENT 4

Method

Materials. Experiment 4 was a replication of Experiment 1 in every respect, except that the items were those recorded for Experiment 3. Instead of 32 targets, as in Experiment 3, there were only 30 targets, as in Experiment 1 (the two additional targets used in Experiment 3 were omitted). Since the preceding context syllables of four targets had been changed in Experiment 3, these items were not completely identical to those used in Experiment 1. During the recording of the Experiment 3 materials, all 80 fillers from both Experiments 1 and 2 were rerecorded. Those from Experiment 1 were used here (half of these were used in Experiment 3). The materials in Experiment 4 therefore closely matched those of Experiment 1, but were based on the new recording.

Subjects and procedure. Twenty-four subjects, none of whom had participated in the previous experiments, took part. Twelve subjects heard each list. As in Experiment 1, subjects were instructed that they would hear nonsense items that could contain finally embedded real words and were asked to press the response key if they heard a nonsense word ending with a real word.

Results and Discussion

All RTs were again originally measured from string onsets. These raw RTs were corrected, by subtracting total string duration, so as to measure effectively from target word offset. In the analysis of the subjects' spoken responses, all responses were found to be correct. Outlying responses were treated as errors (11% of the data). All subjects detected at least 50% of the targets they heard. All targets were detected at least 50% of the time, except one, which was missed by all subjects who heard it in its disharmonious context (*tupa* in *tytupa;* this item was also always missed in Experiment 3). This item was removed from the analysis, leaving 14 back-harmony and all 15 front-harmony targets. The mean RTs and error rates are shown in Table 4.

TABLE 4

MEAN REACTION TIME (RT, IN MS) MEASURED FROM TARGET WORD OFFSET AND MEAN PERCENTAGE ERROR RATES (ERR) IN EXPERIMENT 4

	Target harmony class					
	Back		Front		Overall	
Context	RT	Err	RT	Err	Rt	Err
Harmonious	736	8%	777	13%	757	11%
Disharmonious	693	11%	570	5%	631	8%

In the RT analysis there was a reliable harmony effect. Targets in disharmonious strings were detected, on average, 126 ms faster than targets in harmonious strings: $F1(1,22) = 16.44$, $p < .001$; $F2(1,27) = 11.06$, $p < .005$. Again, this main effect interacted with harmony class. The harmony effect was larger for targets with front vowels (207 ms, on average) than for targets with back vowels (43 ms, on average): $F1(1,22) = 22.47$, $p < .001$; $F2(1,27) = 8.44$, $p < .01$. Pairwise comparisons showed that the effect for back vowel targets was not reliable though that for front vowel targets was: $t1(23) = 5.84$, $p < .001$; $t2(14) = 5.08$, $p < .001$. No other effects were reliable.

In the analysis of error rates, there was no overall harmony effect, but there was a significant interaction of vowel harmony and target harmony class: $F1(1,22) = 14.68$, $p < .001$; $F2(1,27) = 8.04$, $p < .01$. Front-vowel targets were spotted more accurately in disharmonious than in harmonious strings (8%, on average): $t1(23) = 3.61$, $p < .005$; $t2(14) = 4.68$, $p < .001$. Although back-vowel targets were spotted more accurately in harmonious than in disharmonious strings (3%, on average), this difference was not significant. No other effects were significant in the error analysis.

The target durations were reanalyzed, based on the targets analyzed in Experiment 4 (two fewer items were tested and more items reached the 50% criterion than in Experiment 3). A very similar pattern to that found for the Experiment 3 set was nevertheless found: Targets in disharmonious contexts (384 ms, on average) were significantly longer than those in harmonious contexts (369 ms, on average): $F(1,27) = 10.51$, $p < .005$. This was due entirely to the front-harmony targets (mean durations: 392 ms, disharmonious; 364 ms, harmonious) and not to back-harmony targets (mean durations: 376 ms, disharmonious; 374 ms, harmonious), as the interaction showed: $F(1,27) = 7.16$, $p < .05$. Correlational analyses were performed. For back-vowel targets, there were no reliable correlations of target duration, with either RT or error rate. For front-vowel targets, there were also no reliable correlations of target duration with error rate. But there was a reliable negative correlation of front-harmony target duration with RT (faster responses to longer words) for targets in harmonious contexts ($r(14) = -.60$, $p < .05$), but not for targets in disharmonious contexts. Although target duration may have contributed to speed of response within one cell of the design, it cannot account for the overall pattern of results.

A lexical competition analysis was again performed. Responses to targets that in harmonious contexts had competitors consistent with the first two syllables of the string were separated from responses to other targets. As in the analysis of final targets in Experiment 3, there was an interaction of the competition factor with the harmony effect that was only significant by subjects: $F1(1,22) = 6.47$, $p < .05$; $F2(1,25) = 2.67$, $p > .1$. Unlike in Experiment 3, the interaction was in the direction predicted by competition: The difference between harmonious and disharmonious contexts was larger for targets with competitors in the harmonious context (162 ms, on average) than for targets with no competitors in either context (40 ms, on average). The interaction was not significant by subjects or items in the error analysis (a mean difference of 3% for targets with competitors in the harmonious context, and of 4% for targets without competitors). In both Experiments 1 and 4, therefore, there was a nonsignificant trend, suggesting that competitors (such as *kupari* in *kupalo*) can make detection of the target (*palo*) more difficult. This trend was reversed, and again nonsignificant, in Experiment 3. The harmony effect cannot be due to competition between candidate words in harmonious strings.

The results of Experiment 4 broadly replicate those for the final targets in Experiment 3. There is a powerful harmony mismatch effect for targets with front vowels, but a fragile and statistically less reliable effect for targets with back vowels. This is in contrast to Experiment 1, where although the effect was smaller for the back-harmony targets, it was still reliable. The weakness of the effect in the back-harmony words in Experiments 3 and 4 may therefore be in part due to the recording used. Nevertheless, it appears to be a robust finding that the harmony effect is much stronger for front-harmony than for back-harmony targets. For front-harmony targets, at least, a harmony mismatch provides a segmentation cue that assists in word recognition.

EXPERIMENT 5

A final control experiment was performed to deal with a further important concern. It remains possible that the harmony effects obtained are not in fact due to the presence of a harmony mismatch between the preceding context syllable and the target word. Instead, the effects could simply be due to acoustic differences between targets in the two harmony contexts: Perhaps the words spoken in the context of disharmonious syllables were easier to recognize than those spoken in the context of a harmonious syllable simply because those in a disharmonious context were articulated more clearly. Experiment 5 addressed this issue directly. The target words used in Experiment 3 were excised from their contexts and presented to listeners in isolation. Listeners, instead of performing word spotting, were asked to do lexical decision, that is, to press a button every time they detected a real word in a list of words and nonwords. This task is very similar to word spotting and has been used previously (Cutler & Norris, 1988) to address concerns about the acoustic equivalence of word-spotting targets over contexts. If there is some acoustic feature of the words taken from disharmonious contexts, which made them easier to detect than those in harmonious contexts in word spotting, they should likewise be easier to detect in lexical decision. The failure to find such a difference would instead suggest that it is the disharmony between target word and context syllable that produces the harmony mismatch effect.

A related issue to be addressed in Experiment 5 concerns the prosodic patterns of the nonsense strings used in the previous experiments. Finnish is a fixed-stress language, with word-initial syllables as the designated location for word-level stress. As noted above, all trisyllabic strings were produced with the canonical prosody, hence, with stress on the first syllable of the string. Thus, the targets in initial position (with following context, as in Experiments 2 and 3) could perhaps have had a better approximation to their normal stress pattern than those in final position (with preceding context, as in Experiments 1, 3, and 4). It is therefore possible that the harmony effects obtained (i.e., those for words with preceding contexts) are limited to a situation in which the target word has an abnormal stress pattern. However, in the situation where the words had the canonical stress pattern (i.e., words with a following context), no harmony effects were observed, perhaps because normal prosody is sufficient for lexical access, and there is no further role for harmony mismatch to play. If this argument were correct, target words excised from initial positions would be easier to recognize (because of their normal stress patterns) than those excised from final positions (because of a perceived lack of stress on their initial syllables).

Method

Materials. The materials were based on those used in Experiment 3. Using the EASYST speech editor, the 32 target words were excised from each of the four context syllables with which they had been presented in Experiment 3, and for each of the 80 filler items, the harmonious portion was excised from its context syllable, generating 80 bisyllabic nonwords. Cuts were made at zero-crossings closest to the offset of periodic energy associated with either the first vowel of the original trisyllabic string (for targets that had been preceded by a context syllable, such as *palo* in *kupalo*, and for similar filler nonwords) or the second vowel of the original trisyllabic string (for targets that had been followed by a context syllable, such as *palo* in *paloku*, and for similar filler nonwords).

Items were presented in four different lists, each containing all 80 nonwords (from the original fillers) and 32 target words (extracted from one of the four embedding conditions). The counterbalancing and presentation order was identical with Experiment 3. Therefore, in effect, the same four lists were used but, in contrast with the earlier experiment, each item was bisyllabic, the context syllables of both words and nonwords having been removed.

Subjects. Thirty-two student volunteers took part. None of these subjects had participated in the previous experiments. Eight subjects heard each list.

Procedure. The experiment was modeled as closely as possible on Experiment 3. Subjects were told they would hear a list of words and nonwords and were asked to press the response key whenever they heard a real word and then to say aloud the word that they had detected.

Results and Discussion

As in the other experiments, the subjects' spoken responses were analyzed. There were no incorrect spoken responses: When subjects pressed the response key, they always then said the target word. Since all RTs were originally measured from target word onset, they were adjusted so as to measure from word offset using the measurements obtained in Experiment 3. Outlying responses were treated as errors (12% of the data; most of these were extremely fast responses, some made before word offset). No subject failed the exclusion criterion. This criterion was not applied to items, however. Instead, in order to make the most direct comparison with Experiment 3, the seven items that were taken out of that analysis were also excluded here. The mean RTs and error rates are shown in Table 5.

ANOVAs on the RTs showed that, on average, targets that had appeared in final position were detected 51 ms more rapidly than those that had appeared in initial position ($F1(1,28)$

TABLE 5

MEAN REACTION TIME (RT, IN MS) MEASURED FROM TARGET WORD OFFSET AND MEAN PERCENTAGE ERROR RATES (ERR) IN EXPERIMENT 5

	Target harmony class					
	Back		Front		Overall	
	Targets excised from initial position					
Context	RT	Err	RT	Err	RT	Err
Harmonious	447	8%	404	2%	426	5%
Disharmonious	430	15%	400	8%	415	12%
	Targets excised from final position					
Context	RT	Err	RT	Err	RT	Err
Harmonious	368	12%	380	15%	374	14%
Disharmonious	375	9%	361	19%	368	14%

Note. Context (Harmonious/Disharmonious) refers to the contexts from which the target words were excised.

$= 29.0$, $p < .001$; $F2(1,23) = 4.24$, $p = .05$). There was no harmony effect ($F1$ and $F2 < 1$); targets that had appeared in harmonious contexts were detected no more or less rapidly (mean 400 ms) than those that had appeared in disharmonious contexts (mean 391 ms). There was also an effect of the harmony class of the target words: Front-harmony words (like *hymy*) were detected more rapidly (mean 386 ms) than back-harmony words (like *palo;* mean 405 ms). But this effect was only significant in the subjects analysis: $F1(1,28) = 6.95$, $p < .05$; $F2(1,23) = 1.22$, $p > .2$. There were no reliable interactions between the position, harmony, and harmony class factors, though the interaction of the harmony class of the targets with the position in which the target had been embedded was significant by subjects: $F1(1,28) = 4.40$, $p < .05$; $F2(1,23) = 1.64$, $p > .2$. The overall advantage for front over back-harmony words was due to responses to targets that had appeared in initial positions (a 36-ms difference, on average), not to responses to targets that had appeared in final positions (a 1-ms difference, on average).

Error rates were quite high, either because the items were excised from contexts rather than natural tokens spoken in isolation or because some responses were extremely fast (made before word offset) and were counted as errors. ANOVAs on the error rates produced no reliable effects. Several effects were significant in the subjects analysis, but no effects were significant in the items analysis. There was an effect of position, with targets that had been in final positions detected less accurately (14% misses, on average) than those that had been in initial positions (8% misses, on average; $F1(1,28) = 7.36$, $p < .05$, but $F2(1,23) = 1.32$, $p > .25$). There was also a small inverse harmony effect, with rather more errors (mean 13%) to targets that had been in disharmonious contexts than to those that had been in harmonious contexts (mean 9%), but this effect was not significant: $F1(1,28) = 2.88$, $p = .1$, $F2(1,23) = 1.24$, $p > .25$). Although there was no main effect of the harmony class of the targets (both $F1$ and $F2 < 1$), there was an interaction of this factor with target position: For words that had been presented in initial position, those with back vowels were detected less accurately (12% misses, on average) than those with front vowels (5%, on average), while for words that had been presented in final position, those with back vowels were detected more accurately (11% misses, on average) than those with front vowels (17%, on average). But again this effect was only significant in the subjects analysis: $F1(1,28) = 9.51$, $p < .005$; $F2(1,23) = 2.86$, $p > .1$.[2]

These results suggest strongly that the harmony effects found in Experiments 3 and 4 were due to the harmonious or disharmonious contexts in which the target words were presented, and were not due to any acoustic differences between the targets from different contexts. In fact, the only indication of any difference between items due to the harmony context from which they had been excised was in the error analysis, where targets from disharmonious contexts were detected less accurately than those from harmonious contexts (though this was not significant). This effect is, of course, in the opposite direction of that found in the earlier experiments, where targets in disharmonious strings were detected faster (and somewhat more accurately) than those in harmonious strings.

These results are also relevant to concerns about the prosodic patterns of the trisyllabic strings used in the earlier experiments. If words perceived as stressed on their initial syllable were to be easier to recognize than those perceived to have no stress on their initial syllable, then the targets excised from initial position should have been easier to recognize than those excised from final position. In fact, exactly the opposite occurred: Targets from final positions were detected more rapidly than those from initial positions. This result suggests that word stress may not play an important role in the recognition of Finnish speech. The words without stress on their first syllables were easier to recognize than those with stress on their first syllable. It is thus very unlikely that the harmony mismatch effects obtained in Experiments 1, 3, and 4 emerged because of the absence of canonical stress cues. Instead, the results of Experiment 5 suggest that the harmony mismatch effect in our experiments was not determined by the prosody of the trisyllabic string and, hence, that the effect is generalizable to normal fluent Finnish speech recognition.

GENERAL DISCUSSION

In five experiments we have examined the usefulness of Finnish vowel harmony patterns in the process of lexical segmentation. Listeners were faster at detecting words that were preceded by disharmonious contexts, which signal that context and target word are not part of the same lexical entity, than words preceded by harmonious contexts, which are consistent with a single lexical entity embracing both context and target word. This effect was not a function of the competition from other word candidates in harmonious but not in disharmonious strings, since it appeared even when there were no cross-boundary competitors in the harmonious strings. The harmony mismatch effect was significantly stronger for words containing vowels in the front harmony class than for words containing vowels in the back harmony class. Context that followed the target word had no effect upon word detection as a function of harmonic match versus mismatch, for words of either harmony class. Experiment 5 showed that the mismatch effect was not a consequence of acoustic differences between the target words used in each context. It also showed that the fact that the harmony effect was limited to targets with preceding contexts was also not a consequence of acoustic differences (such as those due to stress patterns) between the targets in final and initial positions.

It appears, therefore, that vowel harmony mismatch is of value to listeners during on-line processing of speech, in that it facilitates detection of the beginning of a new word. In a language with word-level vowel harmony, the correspondence between harmony mismatch and the presence of a word boundary is of functional importance to the processing carried out by native listeners.

In fact, the word is effectively the default do-

---

[2] An analysis of all items (i.e., including those items that had failed the Experiment 3 criterion) gave the same pattern of results: There were no reliable effects of the harmony context in which words had originally occurred, and responses to targets from final positions were reliably faster than those to targets from initial positions. Analyses in which no outliers were excluded (i.e., the fast responses, made before word offset were included) also revealed this pattern. In both RT and error analyses, there were no reliable effects of the harmony context from which targets had been excised. The only reliable effect was again that responses were faster to words taken from final positions than to those taken from initial positions.

main of vowel harmony. Although there are a few languages in which vowel harmony spreads across word boundaries (e.g., Somali) or is restricted to a subword level (e.g., Arabic; Vago, 1994), in the vast majority of languages with vowel harmony, the harmony domain is the word plus its affixes. Interestingly, the word as it applies here is not the word as it functions as a syntactic unit, but might perhaps better be termed the lexically represented unit plus its affixes, since as van der Hulst and van de Weijer (1995) point out, compound words usually constitute as many harmony domains as they contain stems. Thus, the natural domain of this phenomenon across languages appears to coincide with the level at which listeners need to segment spoken language for lexical access, and consequently it is perhaps hardly surprising that listeners exploit it to this end. This is not to claim that the presence of vowel harmony in a language like Finnish is per se motivated by its value as a segmentation cue. It can, at best, be merely a partial cue, since only harmony mismatch provides relevant information. There is no bar to the successive occurrence in an utterance of words containing vowels of the same harmony class, with the result that a harmony match is completely uninformative; no harmony pattern can ever signal that a boundary has not occurred (see Suomi, 1983, for further discussion). Nevertheless, our experiments indicate that listeners can indeed efficiently extract what information the harmony patterns do provide.

Vowel harmony offers, therefore, a further (language-specific) regularity that listeners can use to overcome the segmentation problem, that is, the absence of reliable and robust cues to word boundaries in continuous speech. It joins language-specific metrical structure (Cutler et al., 1986; Cutler & Norris, 1988; Otake et al., 1993) and language-specific phonotactic constraints (McQueen & Cox, 1995) as sources of relevant information that can apparently be exploited by listeners in on-line speech processing.

One aspect of our results that was not specifically predicted but that has potentially interesting consequences is the asymmetry in the effect for the two harmony classes of Finnish vowels. Although we observed response facilitation due to harmony mismatch for words with vowels in the back harmony class, the effect here was small and statistically weak compared with the very robust effect that was consistently found in Experiments 1, 3, and 4 for words with vowels in the front harmony class. The latter set, as it happens, may be perceptually disadvantaged. The front harmony vowels (/y ø æ/) are realized closer to the center of the vowel space than either the back harmony vowels (/u o a/) or the so-called neutral vowels (/i e/), which are all peripheral vowels. Furthermore, the front harmony vowels occupy positions in the vowel space closer to the neutral vowels than to the back vowels. The nonperipheral front harmony vowels may therefore in general be more confusible than the peripheral back harmony vowels, and they are likely to be more easily confusible with the neutral vowels than the back vowels are with either the neutral or the front harmony vowels (Suomi, 1983, 1984).

Moreover, front-class vowels have a much lower frequency of occurrence than back-class vowels. In a corpus containing nearly 2 million vowel tokens (Pääkkönen, 1990), approximately 46% of the tokens were from the back harmony class, 39% from the neutral class, and only 15% from the front harmony class. It could therefore be argued that words containing front-class vowels are simply harder to detect than words containing back-class vowels, so that the former benefit more than the latter from a boundary cue that allows efficient initiation of lexical access.

We observed, in fact, no such overall difference in recognizability. Words containing front-class vowels were, if anything, somewhat easier to detect, at least in four of the five experiments reported here. Nevertheless, the asymmetry in the size of the harmony mismatch advantage for the two harmony classes may indeed reflect differences in perceptibility between the two vowel harmony classes. Suppose that such a perceptibility disadvantage is exacerbated for a single front-class vowel in a context containing otherwise only back-class vowels [note that such a case virtually never occurs in Finnish, and Suomi (1983, 1984) has argued that the vowel harmony system itself may be motivated by a pressure to ensure better perceptibility conditions for front-class vowels]. These conditions occur in just one of our four possible stimuli, namely, a back-class word with a disharmonious front-class context syllable (such as *kypalo*). It could be the case that on a small proportion of occasions listeners misidentified the vowel in the context syllable of such items and erroneously perceived it as, for example, /i/ (one of the two most frequent vowels in Finnish, in fact, accounting for 22% of the tokens in the corpus described above). Because /i/ is a neutral vowel and can occur with vowels of both other classes, such a misidentification would convert a disharmonious context to a harmonious one, so that there would be no mismatch and the number of mismatching contexts would thus effectively be lowered for any subject who made this error. We have no way of testing directly whether such misperceptions occurred, but their effect would certainly be to weaken the mismatch effect for the back-harmony words in comparison with the front-harmony words. And note that even if a misperception of /y/ as /i/ were to occur in the harmonious context syllable of a front-harmony word, it would have no effect on the harmony manipulation, since with a neutral vowel the context would remain harmonious.

As long as vowels are perceived to be from mismatching classes, then the vowel harmony system of Finnish provides information that listeners can exploit in locating word onsets. Our experiments have shown that the information is useful, and used, even though it is partial. As we observed in the introduction, vowel harmony allows some exceptions; in modern Finnish, loan words from other languages that violate Finnish vowel harmony constraints are no longer regularized to conform to the native pattern but maintain their mismatching structure. Thus the word *labyrintti* (borrowed into Finnish from Swedish) has a vowel of the back harmony class in the first syllable, and a vowel of the front harmony class in the second syllable, but its recognition would presumably not be facilitated by the postulation of a lexical boundary between the first and the second syllable. Moreover, vowel harmony information can never rule out a word boundary, and many word boundaries will not be cued by harmonic mismatch simply because two successive words belong to the same harmony class or because a syllable adjacent to a word boundary contains a neutral vowel. However, cues to word boundaries do not have to be fully deterministic for listeners to make use of them. Metrical patterns, too, provide only partial and imperfect segmentation cues. Although most lexical words in spoken English begin with strong syllables (Cutler & Carter, 1987), some do not, and although most strong syllables are word initial, some, again, are not. Despite this, English listeners use the occurrence of a strong syllable as a useful heuristic cue to the onset of a new word (Cutler & Butterfield, 1992; Cutler & Norris, 1988), just as Finnish listeners treat a vowel harmony mismatch as highly probably a boundary correlate.

Further, just as metrical information cues word onsets more effectively than word offsets, so too is vowel harmony information used only in the determination of word beginnings. Cutler and Norris (1988) found that the detection of a CVCC word such as *mint* in a sequence of two strong syllables (*mintayf*) was inhibited. They interpreted this finding as evidence that segmentation of the string at the onset of the second strong syllable (*-tayf*) had interfered with the detection of the word which was in fact embedded across this boundary. Detection of a CVC word in a similar sequence (e.g., *thin* in *thintayf*) was, however, not facilitated; that is, the fact that the segmentation would indicate where *thin* ended did not assist listeners in recognizing *thin.* Norris et al. (1995) found that the presence of many potential competitor words for a second syllable could facilitate detection of a CVC word, but this effect was much weaker than the inhibition of CVCC word detection by the presence of a misleading onset cue as in *mintayf.* Cutler and Norris (1988) argued that the process of word recognition in continuous speech contexts is facilitated by information about where to start lexical access attempts; information about where words end is, however, redundant since the same information is, of course, encoded in the lexical entry. Access does not require the complete form to be available but can begin in a continuous manner from the point at which an onset is determined. In an exactly parallel fashion, therefore, a vowel harmony mismatch that signals that a word has ended appears to be of no use in word spotting (Experiments 2 and 3), although a mismatch that cues a word onset is exploited to great effect (Experiments 1, 3, and 4).[3] Thus, the exploitation of vowel harmony patterns in speech segmentation appears to be very similar in kind to the exploitation of metrical patterns.

The metrical pattern found in the English opposition of strong and weak syllables does not occur in Finnish, however. Finnish is a fixed-stress language, with word-initial syllables marked as the location for stress. However, it is important to note that this does not imply an acoustic difference between stressed and unstressed syllables of the type found in English. Finnish has no vowel reduction, so that there are no vowel quality differences between stressed and unstressed syllables such as typically occur in English. Finnish also has phonemic vowel quantity distinctions, which are orthogonal to stress placement, so that the duration differences between stressed and unstressed syllables which are found in a language like English also do not occur in Finnish; in a word such as *vapaaseen* (''into a free''), with a short vowel in the initial syllable and a long vowel in the second and third syllables, the initial syllable, although the designated stressed syllable, will be the shortest of the three. Differences between syllables in amplitude and fundamental frequency are also not large, although, as in most languages of the world, there will be downdrift across the word in both these dimensions. In three-syllable sequences, such as were used in our experiments, there will also be no internal foot boundaries which could affect perceived prominence.

Thus it is doubtful whether stress as the term is applied to Finnish should be taken to have processing implications. In fact, it may be argued that the stress patterns of our stimulus words satisfied the canonical word prosody of Finnish both with preceding and with following context. If the canonical pattern of a three-syllable word is taken to be downdrift in both amplitude and fundamental frequency contour across the word, with no durational differences between syllables other than those resulting from segmental structure, then our words had this structure irrespective of context position. Indeed, in the judgment of the first author of this paper, the words of Experiment 5 which had been removed from preceding contexts did not sound prosodically abnormal, and the behavior of the subjects in that experiment is fully consistent with this. Prosodic structure of this kind does not offer a segmentation cue to listeners in the way that English metrical structure does; vowel harmony, on the other hand, is in a position to provide Finnish listeners with such a cue. As our results showed, listeners were able to use the cue it provided. Any effects of word prosody consequent upon position in the string should have been exactly the same in both harmonious and disharmonious contexts; and Experiment 5 confirmed that there were in fact no such differences *between* the contexts. Thus the strong and consistent difference that we observed between harmonious and disharmonious contexts can only be interpreted as confirming the suggestion originally made by Trubetzkoy (1939)—that vowel harmony can function as a segmentation cue.

The use of segmentation cues in continuous speech recognition has been successfully modeled in computer simulations. The competition mechanism in the Shortlist model can be enriched, for example, with a metrical segmentation process, such that the model accurately captures the patterns observed in human data (Norris et al. 1995; Norris, McQueen, Cutler, & Butterfield, 1996). These simula-

---

[3] Following contexts may have been of little value for listeners not only because word offsets can in general be established from lexical information, but also because the particular target words used in these experiments tended to be unique at their offsets (only 7 words became unique after their final vowels). Word offsets could thus be established both from the lexicon and, for most words, from the segmental information in the input. Words that do not become unique until after their offsets may be more likely to show effects of vowel harmony mismatch from following contexts. But post hoc analyses on the Experiment 2 data (where UP effects should have been strongest because of fixed-target location) failed to show any reliable differences between words that were or were not unique by offset.

tions have been based on data from stress-timed languages (English and Dutch), where information about the location of strong syllable onsets, provided in the input, was used to bias the competition process. There is no reason, however, why the same approach could not be taken with other languages (Cutler, Norris, & McQueen, 1996). In a version of Shortlist for processing French, for example, information about the location of all syllable boundaries could be used to influence the competition process. Language-specific metrical segmentation routines can all be implemented in a universal processing model, where the core mechanism is that of competition between candidate words.

As with segmentation cued by metrical structure, it is likewise possible to implement segmentation cued by phonotactics or by silence in the Shortlist model (McQueen & Cox, 1995; Norris et al., 1996). Our present experiments have shown that vowel harmony is one further segmentation cue that Finnish listeners can use. In principle, therefore, it would be possible to use vowel harmony mismatch to bias the competition process in Shortlist. This implementation, however, is not feasible until a machine-readable Finnish lexicon becomes available. Nevertheless, the present results are consistent with the view of continuous speech recognition instantiated in the Shortlist model. Vowel harmony provides yet another means by which clear boundaries may be signaled in the speech signal.

APPENDIX: EXPERIMENTAL MATERIALS

All words were used in all experiments, except maku and jyvä, which were only used in Experiments 3 and 5. Each word is listed with preceding (CV-) and following (-CV), harmonious and disharmonious context syllables. Some items required more than one context syllable in initial position (see text for details).

Harmony class

Harmonious context			Disharmonious context			
CV-	Word	-CV	CV-	Word	-CV	Gloss
			Back			
ku	palo	ku	ky	palo	ky	fire
ka	kuja	ka	kä	kuja	kä	alley
pu/po	lato	pu	py/pö	lato	py	barn
tu	haka	po	ty	haka	pö	hook
to	luku	to	tö	luku	tö	number
pu	juna	po	py	juna	pö	train
po	sopu	ta	pö	sopu	tä	agreement
ku	romu	ku	ky	romu	ky	trash
po	kuva	po	pö	kuva	pö	picture
po	muna	po	pö	muna	pö	egg
to	latu	to	tö	latu	tö	track
ta	raju	pu	tä	raju	py	rash
pu/tu	tupa	tu	py/ty	tupa	ty	cottage
ku	koru	ku	ky	koru	ky	piece of jewelery
tu	napa	tu	ty	napa	ty	navel
ta	maku	ta	tä	maku	tä	taste

APPENDIX—*Continued*

Harmony class

Harmonious context			Disharmonious context			
CV-	Word	-CV	CV-	Word	-CV	Gloss
			Front			
ty	kynä	pö	tu	kynä	po	pen
py/tä	näkö	tä	pu/ta	näkö	ta	sight
kä	pöly	kä	ka	pöly	ka	dust
ky	sävy	ky	ku	sävy	ku	shade
pö/ty	hätä	ty	po/tu	hätä	tu	emergency
ky	pyry	ky	ku	pyry	ku	snowfall
ty	kyky	ty	tu	kyky	tu	ability
pö	käry	py	po	käry	pu	odour
tö	häkä	tö	to	häkä	to	carbon monoxide
py	hymy	py	pu	hymy	pu	smile
pö	läjä	pö	po	läjä	po	heap
tö	käpy	tö	to	käpy	to	pine cone
ky	rysä	ky	ku	rysä	ku	trap
pö	syvä	pö	po	syvä	po	deep
tä	tyly	pö	ta	tyly	po	harsh
tö	jyvä	tö	to	jyvä	to	grain

REFERENCES

CHRISTIE, W. M. (1974). Some cues for syllable juncture perception in English. *Journal of the Acoustical Society of America,* **55,** 819–821.

CHURCH, K. W. (1987). Phonological parsing and lexical retrieval. *Cognition,* **25,** 53–69.

CLUFF, M. S., & LUCE, P. A. (1990). Similarity neighborhoods of spoken two-syllable words: Retroactive effects on multiple activation. *Journal of Experimental Psychology: Human Perception and Performance,* **16,** 551–563.

CUTLER, A., & BUTTERFIELD, S. (1992). Rhythmic cues to speech segmentation: Evidence from juncture misperception. *Journal of Memory and Language,* **31,** 218–236.

CUTLER, A., & CARTER, D. M. (1987). The predominance of strong initial syllables in the English vocabulary. *Computer Speech and Language,* **2,** 133–142.

CUTLER, A., MEHLER, J., NORRIS, D. G., & SEGUÍ, J. (1986). The syllable's differing role in the segmentation of French and English. *Journal of Memory and Language,* **25,** 385–400.

CUTLER, A., MEHLER, J., NORRIS, D. G., & SEGUÍ, J. (1992). The monolingual nature of speech segmentation by bilinguals. *Cognitive Psychology,* **24,** 381–410.

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.

CUTLER, A., NORRIS, D., & MCQUEEN, J. (1996). Lexical access in continuous speech: Language-specific realisations of a universal model. In T. Otake & A. Cutler (Eds.), *Phonological structure and language processing: Cross-linguistic studies* (pp. 227–242). Berlin: Mouton de Gruyter.

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

GOLDINGER, S. D., LUCE, P. A., PISONI, D. B., & MARCARIO, J. K. (1992). Form-based priming in spoken word recognition: The roles of competition and bias. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* **18,** 1211–1238.

GOW, D. W., & GORDON, P. C. (1995). Lexical and prelexical influences on word segmentation: Evidence from priming. *Journal of Experimental Psychology: Human Perception and Performance,* **21,** 344–359.

VAN DER HULST, H., & VAN DE WEIJER, J. (1995). Vowel harmony. In J. A. Goldsmith (Ed.), *The handbook of phonological theory* (pp. 497–534). Oxford: Blackwell.

KARLSSON, F. (1983). *Suomen kielen äänne- ja muotorakenne (Finnish phonological and morphological structure).* Helsinki: WSOY.

LEHISTE, I. (1972). The timing of utterances and linguistic boundaries. *Journal of the Acoustical Society of America,* **51,** 2018–2024.

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

MCQUEEN, J. M., & COX, E. (1995). The use of phonotactic constraints in the segmentation of Dutch. In *Proceedings of Eurospeech '95,* Vol. 3, 1707–1710, Madrid, Spain.

MCQUEEN, J. M., CUTLER, A., BRISCOE, T., & NORRIS, D. (1995). Models of continuous speech recognition and the contents of the vocabulary. *Language and Cognitive Processes,* **10,** 309–331.

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.

MEHLER, J., DOMMERGUES, J.-Y., FRAUENFELDER, U. H., & SEGUÍ, J. (1981). The syllable's role in speech segmentation. *Journal of Verbal Learning and Verbal Behavior,* **20,** 298–305.

MEISTER, E., & SUOMI, K. (1993). EASYST—A computer system for reaction time measurements. In A. Iivonen & E. Aulanko (Eds.), *Fonetiikan päivät—Helsinki 1992. Papers from the 17th meeting of Finnish Phoneticians,* pp. 131–143. Helsinki: Department of Phonetics, University of Helsinki.

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

NORRIS, D. G. (1990). A dynamic-net model of human speech recognition. In G. T. M. Altmann (Ed.), *Cognitive models of speech processing: Psycholinguistic and computational perspectives* (pp. 87–104). Cambridge, MA: MIT Press.

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

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

NORRIS, D., MCQUEEN, J. M., CUTLER, A., & BUTTERFIELD, S. (1996). *The possible-word constraint in the segmentation of continuous speech.* Manuscript submitted for publication.

OTAKE, T., HATANO, G., CUTLER, A., & MEHLER, J. (1993). Mora or syllable? Speech segmentation in Japanese. *Journal of Memory and Language,* **32,** 258–278.

PÄÄKKÖNEN, M. (1990). *Grafeemit ja konteksti: Tilastotietoja suomen yleiskielen kirjaimistosta* (Graphemes and context: Statistical counts of the letter frequencies in Standard Finnish). Vaasa: Suomalaisen kirjallisuuden seura.

PALLIER, C., SEBASTIÁN-GALLÉS, N., FELGUERA, T., CHRISTOPHE, A., & MEHLER, J. (1993). Attentional allocation within the syllable structure of spoken words. *Journal of Memory and Language,* **32,** 373–389.

QUENÉ, H. (1987). Perceptual relevance of acoustical word boundary markers. *Proceedings of the Eleventh International Congress of Phonetic Sciences,* Tallinn; Vol. 6, pp. 79–82.

SEBASTIÁN-GALLÉS, N., DUPOUX, E., SEGUÍ, J., & MEHLER, J. (1992). Contrasting syllabic effects in Catalan and Spanish. *Journal of Memory and Language,* **31,** 18–32.

SHIMIZU, K., & DANTSUJI, M. (1980). A study on perception of internal juncture in Japanese. *Studia Phonologica,* **14,** 1–15.

SUOMI, K. (1983). Palatal vowel harmony: A perceptually motivated phenomenon? *Nordic Journal of Linguistics,* **6,** 1–35.

SUOMI, K. (1984). A revised explanation of the causes of palatal vowel harmony based on psychoacoustic spectra. *Nordic Journal of Linguistics,* **7,** 41–61.

TABOSSI, P., BURANI, C., & SCOTT, D. (1995). Word identification in fluent speech. *Journal of Memory and Language,* **34,** 440–467.

TRUBETZKOY, N. S. (1939). Grundzüge der Phonologie. *Travaux du Cercle Linguistique de Prague,* **7.**

VAGO, R. M. (1994). Vowel harmony. In R. E. Asher (Ed.), *The encyclopedia of language and linguistics* (pp. 4954–4958). Oxford: Pergamon.

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., VAN ZON, M., & DE GELDER, B. (1996). Cues to speech segmentation: Evidence from juncture misperceptions and word spotting. *Memory & Cognition,* **24,** 744–755.

YERKEY, P. N., & SAWUSCH, J. R. (1993). The influence of stress, vowel and juncture cues on segmentation. Paper presented at the 126th meeting of the Acoustical Society of America. *Journal of the Acoustical Society of America,* **94,** 1880.

ZWANENBURG, W., OUWENEEL, G. R. E., & LEVELT, W. J. M. (1977). La frontière du mot en français. *Studies in Language,* **1,** 209–221.

(Received January 5, 1996)
(Revision received June 26, 1996)