

Language Specific Cues To Segmentation
of Spoken Words in Finnish
Behavioral and Event-Related Brain Potential Studies
Jyrki Tuomainen

Language Specific Cues To Segmentation of Spoken Words in Finnish Jyrki Tuomainen

LANGUAGE SPECIFIC CUES TO SEGMENTATION OF
SPOKEN WORDS IN FINNISH:
BEHAVIORAL AND EVENT-RELATED BRAIN POTENTIAL STUDIES

JYRKI TUOMAINEN

BRAIN IS RELATED TO BEHAVIOR ($P < .05$)^{*}

^{*}Konstantin K. Zakzanis (1998)
Journal of Clinical and Experimental Neuropsychology, **20**, p. 419.

LANGUAGE SPECIFIC CUES TO SEGMENTATION
OF SPOKEN WORDS IN FINNISH:
BEHAVIORAL AND EVENT-RELATED BRAIN POTENTIAL STUDIES

PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Katholieke Universiteit Brabant,
op gezag van de rector magnificus, prof. dr. F.A. van der Duyn Schouten,
in het openbaar te verdedigen ten overstaan
van een door het college voor promoties aangewezen commissie
in de aula van Universiteit op
vrijdag 18 mei 2001
om 11.15 uur

door

JYRKI JUHANI TUOMAINEN

geboren op 10 april 1958
te Imatra, Finland

Promotores: Prof.dr. B.L.M.F. de Gelder
Dr. J.H.M. Vroomen

ISBN 952-91-3404-5

Printed and bound by Painosalama Oy, Turku, Finland

Cover illustration: 'Aivot myrskyssä' © 2000 Aino Tuomainen

© 2001 Jyrki Tuomainen

LANGUAGE SPECIFIC CUES TO SEGMENTATION OF
SPOKEN WORDS IN FINNISH:
BEHAVIORAL AND EVENT-RELATED BRAIN POTENTIAL STUDIES

Acknowledgments

1. INTRODUCTION	1
2. THE ROLES OF WORD STRESS AND VOWEL HARMONY IN SPEECH SEGMENTATION	19
3. FUNDAMENTAL FREQUENCY IS AN IMPORTANT ACOUSTIC CUE TO WORD BOUNDARIES IN FINNISH	41
4. WORD STRESS IN LEXICAL SEGMENTATION OF SPOKEN FINNISH ...	49
5. TEMPORAL DYNAMICS OF LEXICAL STRESS AND VOWEL HARMONY IN THE LEXICAL SEGMENTATION IN FINNISH: AN EVENT-RELATED BRAIN POTENTIAL STUDY	65
6. GENERAL DISCUSSION AND CONCLUSIONS	107

References

Appendices

Samenvatting

ACKNOWLEDGEMENTS

The main body of the current work was done between 1995-1998 while I was at the Cognitive Neuroscience Lab, Department of Psychology, Tilburg University, Tilburg, The Netherlands. Part of the experiments were run at the Centre for Cognitive Neuroscience (CCN), and Department of Neurology, University of Turku, Finland. Final stages of the writing of the thesis were made possible by a research position at the CCN (fall 1999 and early summer 2000).

I would like to thank the following persons who have greatly helped me with this thesis and otherwise:

- Prof. Beatrice de Gelder, Ph.D., for offering me a position in Tilburg, by creating an inspiring research environment, and supporting all my research activities.
- Jean Vroomen, Ph.D., for scientific advice, teaching me the basics in psycholinguistics, and for being a friend in good and bad (have you already burned up all the firewood?).
- Prof. Anne Cutler, Ph.D., for being the initial force behind my thesis project, and having time for discussions and advice, and also for the membership in the dissertation committee.
- Koen Böcker, Ph.D., for extensive help and advice with ERPs.
- Prof. Shlomo Bentin, Ph.D. (Hebrew University of Jerusalem, Israel), Mireille Besson, Ph.D. (CNRS, Marseille, France), Ton van Boxtel, Ph.D. (KUB, The Netherlands), and professor Kari Suomi, Ph.D. (University of Oulu, Finland) for accepting the invitation to act as members of the dissertation committee.
- Stefan Werner, Ph.D., for co-authorship (Chapter 3) and help in the acoustic analyses of the materials, and also for many good (but short) moments in Joensuu, Finland.
- Prof. Matti Laine, Ph.D., for over 15 years of support in all kinds of activities, numerous advice and help in scientific research, and being a friend.
- Prof. Heikki Hämäläinen, Ph.D., for providing me the opportunity to work as a "lab engineer" during the final stage of writing the thesis, and for discussions and advice.
- Prof. Jussi Niemi, Ph.D., for checking the English of the thesis, and for support throughout the years.
- My family (and especially Minna, Elli, Aino, Lauri, and Heikki) for being there.

I have only mentioned by name those persons who have been around during the time of the current project. However, several other people related to my scientific activities in Tilburg, and in Turku (Centre for Cognitive Neuroscience, Department of Neurology, and Phonetics Lab), and elsewhere have been most helpful in many ways. Accordingly, warm thanks are extended to these anonymous friends and colleagues. Finally, the Academy of Finland is gratefully acknowledged for the financial support during 1995-1998.

Turku, April 10, 2001

CHAPTER 1

INTRODUCTION

Recognition of individual written words is difficult when there are no white spaces between words. sand iti seven mo red if fic ultifthes pace sare
in wrong position is ion sand a sac on sequence the rear en or elia ble cue
stow here on ewordend sand the other be gins.

This example demonstrates the difficulties related to missing and misleading information about word boundaries in visual word recognition. Incidentally, it also captures some aspects of spoken word recognition. The former part of the signal consists only of a string of graphemes without white spaces between words. Nonetheless, it contains all the necessary details so that words can be recognized. The goal, however, is achieved at the expense of time. Word recognition is delayed without additional cues to word boundaries. In the latter part of the example, mislocated cues, the white spaces in wrong positions, make the reader's task more difficult than missing cues.

In the listener's mind, speech consists of a string of separate words. However, in natural speech, unambiguous and reliable cues to word boundaries, such as pauses, are rare. Pauses or silent parts do exist in the acoustic waveform, but often they do not coincide with word boundaries. This is exemplified in Figure 1, which shows the waveform representation of the Finnish sentence "Marja näkee ikkunan läpi äidin" ('Marja sees mother through the window'). The silent parts (marked with small arrows) are of no use in terms of lexical segmentation. They denote the silent portions of plosives, and are not aligned with word onsets. The real word onsets are marked with large arrows. As the figure shows, no clear acoustic indices that would unambiguously provide cues to word boundaries are present. Even though it is possible to segment the acoustic signal based on acoustic criteria, these segments do not seem to correspond in any straightforward manner to linguistic units such as phonemes, syllables or words. Instead, speech input is a string of acoustic events

that are not linearly aligned in terms of phonological segments. (For a discussion of the relationship between speech perception and spoken word recognition, see Nygaard & Pisoni, 1995.) The ease that human listeners show in recognizing words in continuous speech is even more remarkable when one realizes that most words contain embedded words (e.g., *bone* in *trombone*), and some (unintended) words spread over (intended) word boundaries (e.g., in Italian *visite* 'visits' in *visi tediati* 'faces bored'). Nonetheless, the embedded words are practically never consciously recognized during conversation, even though there is evidence that at least some of them are momentarily entertained as possible lexical candidates (e.g., Gow & Gordon, 1995; McQueen, Norris, & Cutler, 1994; Shillcock, 1990; Tabossi, Burani, & Scott, 1995; Vroomen & de Gelder, 1995, 1997).

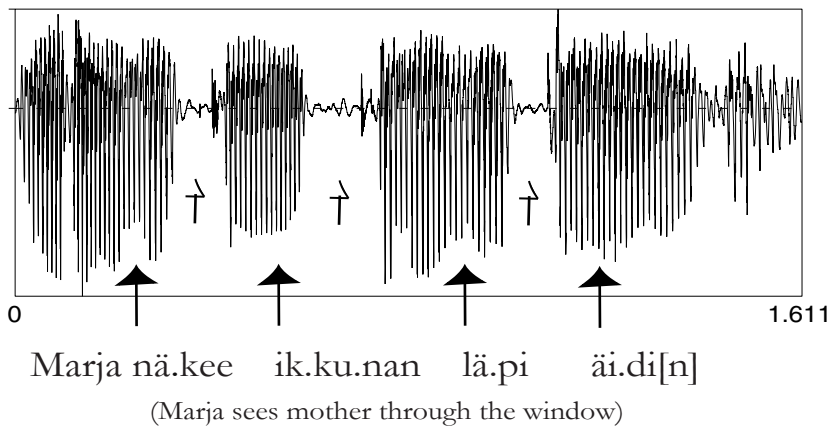


FIGURE 1. Representation of the acoustic waveform of the utterance *Marja näkee ikkunan läpi äidin* ('Marja sees mother through the window'). Large arrows are placed at word onsets, small arrows indicate pauses (or silent parts).

Word boundaries emerge as by-products of word recognition

What makes it so easy for listeners to recognize individual words in continuous speech? Several alternative theories have emerged. A common feature of recent models of spoken word recognition is that lexical processes are mostly responsible for both recognition of words and also for revealing the boundaries. This entails a

logical necessity; once a word is recognized, the onset and offset are known. Examples of models in which word boundaries emerge as by-products of recognition are TRACE (McClelland & Elman, 1986) and Shortlist (Norris, 1994). Although differences exist between these models, a common feature in both of them is that a set of lexical candidates (or hypotheses) are activated by portions of the acoustic input, which in turn compete for selection. As more phonemes are made available, the candidates that do not match the input drop out of the competition. Finally, word recognition occurs when only one candidate has survived. A reasonable assumption is that lexical competition is a universal property of the spoken word recognition mechanism, although direct tests of competition across languages are scarce (e.g., McQueen et al., 1994; Norris et al., 1995; Vroomen & de Gelder, 1995). Given the complexities of the generation the lexical hypotheses, it is probable that additional cues that are present in the speech signal facilitate and speed up recognition of words in continuous speech.

Additional cues make detection of words easier and faster

A growing body of research on spoken language processing indicates that listeners are sensitive to additional cues in the speech signal that help segment speech into meaningful chunks. These cues may be phonetic, such as initial or final lengthening of syllables, aspiration of initial stops, glottal stop, or allophonic variation (e.g., Gow & Gordon, 1995; Gårding, 1967; Klatt, 1975; Lehiste, 1960; 1972; Nakatani & Dukes, 1977), or phonological, such as phonotactic restrictions on what kinds of phoneme sequences can occur at syllable boundaries (McQueen, 1998). Statistical properties of language specific phoneme sequences (that is, transitional probabilities between syllables) have shown to play a role in segmentation (e.g., Saffran et al., 1996). Finally, a special emphasis has been on prosodic features, which create a rhythm that organizes the signal and facilitates lexical segmentation. Cutler and her colleagues have put forth a general hypothesis that the phonological structure of the individual language defines the nature of these cues. Thus, because languages differ in their structure, these cues are language specific. For example, English words consist of strong and weak syllables. A strong syllable contains a full vowel and a weak syllable, usually containing a schwa. Based on observations that about 90% of English words start with strong syllables (Cutler & Carter, 1987), Cutler and Norris (1988) suggested a Metrical Segmentation Strategy (MSS) might be helpful in segmenting continuous English. A viable strategy then could be: "Every time a strong syllable (containing a full vowel) is encountered, assume the existence of a word boundary". Accordingly, segmentation at strong syllables will result either in facilitation or slowing down of RTs depending on whether the candidate is aligned

with the intended word boundary or embedded in another word. For example, the original demonstration by Cutler and Norris (1988) showed that “mint” is more difficult to recognize when it is embedded in [mɪntɛɪf] as opposed to [mɪntəf]. The explanation is that in [mɪntɛɪf] a word boundary is assumed between the two strong syllables so that both [mɪnt] and [tɛɪf] are segmented. It takes extra time to reconstruct [mɪnt] over the segmentation point. Several reports have indicated that both English speaking adults (e.g., Cutler and Norris, 1988), and infants (e.g., Jusczyk, Houston, & Newsome, 1999) are sensitive to the rhythm created by alternating strong and weak syllables, and seem to use MSS in segmentation. In languages that resemble English in terms of the distinction between strong and weak syllables, such as Dutch, the MSS has been shown to operate (Vroomen & de Gelder, 1995; Vroomen, van Zon, & de Gelder, 1996).

The metrical unit that seems to create the rhythm in French is the syllable. Contrary to English, syllable boundaries in French are in general unambiguous (see, however, Content et al., in press, for a demonstration of ambisyllabicity in French). French listeners seem to segment speech input into syllable-sized units (Mehler, Dommergues, Frauenfelder, & Seguí, 1981). In their seminal study, the authors reported that, in a syllable monitoring task, participants were faster in detecting *ba* in *ba.lance* than in *bal.con*. The reverse pattern was found with syllable *bal*. Despite the original claim that syllable is *the* unit that listeners use in lexical access and segmentation, it turned out that the pattern holds only for French and other syllable-timed languages such as Spanish (Pallier, Sebastián-Gallés, Felguera, Christophe, & Mehler, 1993) and Catalan (Sebastián-Gallés, Dupoux, Seguí, & Mehler, 1992). When English participants were presented with the same materials, it did not matter whether the target corresponded or not to the first syllable of the word (Cutler, Mehler, Norris, & Seguí, 1986). However, this is not to say that syllables do not play any role in spoken word recognition and segmentation in English. Norris and his colleagues (e.g., Norris, McQueen, Cutler, & Butterfield, 1997) have put forth a hypothesis what they call a 'Possible Word Constraint' (PWC), which emphasizes the role of the syllable in parsing the spoken input into meaningful units. PWC suggests that listeners parse the input so that all bits and pieces are taken into account. If the parsing of a phoneme string results in a lexical hypothesis and a chunk that is not a possible word in language then that parse is penalized. For example, the target 'apple' is detected faster in *vuffapple* than in *fapple*, because, in *fapple*, 'f' is left out, and it cannot form a word (or a syllable) in English. Thus, PWC emphasizes the role of syllable in word recognition, but the role is different from French. Based on several experiments conducted with listeners with different language backgrounds, PWC currently reads: "Penalize any parse that yields as output a string of phonemes that does not contain a vowel, because it cannot be a

possible word." (Norris, Cutler, McQueen, Butterfield, & Kearns, 2000). PWC has also been incorporated to Shortlist and Merge (Norris, McQueen, & Cutler, 2000).

In Japanese the rhythmic pattern is created by morae. The mora refers to a subsyllabic unit. In a hierarchical structure, it is dominated by the syllable and it in its turn dominates a segment (either a vowel or consonant) or segments. A syllable contains at least one mora. For example, a CV syllable consists of one mora, a CVC syllable consists of two morae (CV and C). Long vowels are considered bimoraic, and nasal consonants and (the first member of a) geminate consonant also contribute to the mora. In some analyses the mora has been considered as both a temporal and a tonal unit; morae affect the rhythm and also the placement of the pitch accent in a word (e.g., Nakano-Madsen, 1992). Japanese listeners seem to be sensitive to the rhythm created by morae, but not syllables, which in turn helps segmentation (e.g., Otake, Hatano, Cutler, & Mehler, 1993).

How do listeners deal with multiple cues to word boundaries?

In face-to-face communication listeners have multiple cues to word boundaries available while they listen to speech. Little attention has been paid on how adult listeners deal with these multiple cues. As a matter of fact, most research reporting on how multiple cues are used has mainly focused on phonetic cues (e.g., Quené, 1993; Yerkey & Sawusch, 1993) or how infants learning their first language deal with multiple phonological cues (Mattys, Jusczyk, Luce, & Morgan, 1999). For example, Quené (1993) investigated how segment durations of the pivotal consonant and the rise time of the postboundary vowel signaled word boundary as a function of the accent level of the syllable. The results showed that listeners utilized duration as a cue to word boundary, but performance was better if the word after the boundary was accented. Accordingly, accent may enhance the perceptual salience of the phonetic cues. Mattys et al. (1999) studied how 9-month-old infants deal with phonotactic regularities and prosodic pattern in segmenting spoken English. The results showed, among other things, that infants were sensitive to probabilistic information yielded by phonotactic sequences. Furthermore, when both cues were present, infants relied more strongly on the prosodic cue (primary or secondary stress) than on the phonotactic one. The authors suggested that as early as at the age of 9 months, children have the capability of integrating multiple cues, which usually improve performance. However, given the preference for the stress cue, Mattys et al. argued that prosodic cues constitute a first-pass strategy that enables English-learning infants to begin to segment content words from continuous speech. Attending to prosodic cues is thus primary, but this strategy is supplemented with

other cues which in most cases increase the success rate of segmentation and diminish false alarms.

Phonological properties of Finnish

Another example of potential cues to word boundaries as suggested by the phonological and prosodic structure of a language comes from Finnish. It is a language spoken by some 5 million speakers mostly in Finland. Finnish belongs to the Finno-Ugric language family and it has a rich morphology with highly complex inflectional system involving both nouns and verbs. In typological classification, Finnish is described as belonging to the syllable-timed languages. Syllable boundaries are clear, and the syllable structure fairly simple. For example, consonant clusters are not permitted at the syllable initial position. Another feature of Finnish is that quantity, the distinction between short and long phonemes, is distinctive. This phonological feature pertains to both vowels and consonants in almost any position of a word. For example, words differing only in phonemic length form the following continuum; *tule* 'come!', (*ei*) *tuule* '(is not) windy', (*ei*) *tuulle* 'is (probably not) windy', *tuulee* 'is windy', *tulee* 'comes', *tullee* 'will probably come', *tuullee* 'it is probably windy'.

Most importantly for lexical segmentation, there are two potential cues to word boundary, one is vowel harmony and the other is lexical stress. Vowel harmony refers to the phonotactic restrictions which define that (native) Finnish word (stems) may only contain vowels from one harmony set (or from the neutral set). Accordingly, vowels /y æ ø/ comprise the front harmony set, and vowels /u o a/ the back harmony set (/i e/ belong to neutral vowels). Words like /katu/ ('street') and /hymy/ ('smile') are viable words, whereas */kætu/ or */katy/ or */humy/ or */hymu/ are not. (Note: in all examples, International Phonetic Alphabet, IPA, is used to indicate phonetic and phonological symbols. In the orthographic notation, Finnish <ä> denotes /æ/ <ö> denotes /ø/). In one sense, vowel harmony can be regarded as a phonological device that increases the coherence of a word form. The other side of the coin is that vowel harmony also provides a means for determining word boundaries (Trubetzkoy, 1958). Thus, when vowels from two opposing harmony categories within a string of phonemes are detected, a word boundary must be present. (For more detailed accounts of the Finnish language, see e.g., Iivonen, 1998; Karlsson, 1987.)

In Finnish lexical (or primary word) stress is fixed on the first syllable of the word. Fixed stress may thus provide a potentially reliable cue to word boundary, because stressed syllables may stand out more prominent than surrounding syllables.

Listeners could start lexical access at every stressed syllable. However, because stress is fixed in Finnish, it supposedly does not have a (major) linguistic function, and in this respect it provides redundant information. However, fixed stress may serve a demarcative function (Hyman, 1977; Trubetzkoy, 1958). In contrast, because of its redundant character in the linguistic sense, fixed lexical stress might not be realized acoustically (Cutler, Dahan, & van Donselaar, 1997) and might be of little use as a cue to a word boundary.

Although systematic studies regarding the acoustic structure of word and sentence level prominence are scarce, in most Finnish textbooks of linguistics and phonetics the acoustic correlates of word stress are described as consisting of fundamental frequency (F_0), amplitude, and duration (e.g., Karlsson, 1983, Iivonen et al., 1987). Fundamental frequency is also the main acoustic correlate of intonation (Iivonen et al., 1987) and F_0 is often (explicitly or implicitly) correlated with word stress in Finnish (Hirst & Di Cristo, 1998). For example, Välimaa-Blum (1993), following the terminology of Pierrehumbert (1980) in which pitch accent is the melodic correlate of stress, suggests that Finnish intonation pattern consists of a sequence of pitch accents and a boundary tone. Pitch accent is realized as a simple F_0 movement occurring at the primary stressed syllable of the word. Specifically, for a neutral declarative sentence, sequence of L+H* pitch accents followed by a boundary tone L% is the default for unmarked word order. (L (Low), H (High), *, and % are transcript characters in the Tone and Break Indices (ToBI) system (Silverman et al., 1992). However, not all first syllables receive the pitch accent. For example, it seems that at least finite verbs in the neutral declarative sentences are unstressed. Thus, word stress will not inform of *all* word boundaries, but will do so on many instances. In a recent study by Iivonen, Niemi, & Paananen (1998) examined the potential use of stress as a cue to word boundaries. They tried to determine the extent to which F_0 peaks in Finnish, English, and German coincide with word stress. They analyzed TV and radio newscasts and counted how often a just noticeable F_0 peak (defined as a difference in one semitone or more when compared with the neighboring syllable) matched a primary stressed syllable. One cannot expect a perfect correlation between F_0 peaks and word stress because stress may not always be acoustically realized. Nevertheless, Iivonen, Niemi, and Paananen found that the majority of Finnish F_0 peaks, 73%, occurred on the primary stressed syllable, while only 42% of the German peaks and 59% of the English peaks represented word stress. Moreover, about 52% of the Finnish word-initial syllables had an F_0 peak. Thus, this phonetic analysis suggests that F_0 peaks are at least partly successful in signaling which syllables receive primary stress, and hence, where a word boundary is located in Finnish speech.

Prosody: a terminological remark

In the paragraphs above, terms such as fundamental frequency, intonation, amplitude, duration, prominence, stress, accent, rhythm etc. are used to describe prosodic features of speech. However, they clearly denote concepts at various levels of description. As such, the diversity of concepts suggests a complex structure, which may be studied separately at each level. Accordingly, the definition of prosody has been elusive so far. Recently, researchers have focussed on how the different levels of description could be organized within a unifying model of prosody (see e.g., Beckman, 1996; Hirst & Di Cristo, 1998; Shattuck-Hufnagel & Turk, 1996 for recent reviews). There are at least two levels of description that should be integrated in a unifying theory of prosody. The first is the physical level of measurable physical parameters, which refers to acoustic parameters such as fundamental frequency, amplitude, duration and spectral properties of a segment. These parameters correlate with different perceived aspects (especially with relative prominence of syllables) of the spoken utterance in a complex way. Second, prosody can be regarded as a phonological organization of segments, which can be parsed into higher-level constituents and to a pattern of relative prominence within these constituents. Recent developments especially within autosegmental phonology and metrical theory (e.g., Liberman and Prince, 1977) (which have been integrated with intonational theory, e.g., Beckman & Pierrehumbert, 1986) suggest that prosody itself is a grammatical (phonological) structure that must be parsed in its own right (e.g. Beckman, 1996; Nespor & Vogel, 1986). The prosodic constituents form a hierarchy spanning from utterance to syllable (and in some theories to subsyllabic units such as mora). Several factors, both linguistic and non-linguistic, influence the speaker's choice of prosody for a given utterance. The prosodic component could be regarded as an 'integrator' of these factors within grammar (Shattuck-Hufnagel & Turk, 1996).

In one sense, prominence is a perceptual property of spoken language. It refers to the perceptual salience of a syllable in relation to the surrounding syllables. Prominent syllables may be acoustically more intense, have a higher pitch, or be longer in duration than less prominent syllables. One type of framework to describe relative prominence levels comes from autosegmental-metrical theory, which assumes four prominence levels of syllables: nuclear accented, accented, heavy, and (reduced or light) syllable (e.g., Beckman, 1996; see also Ladefoged, 1975). According to Beckman (1996) the type of prominence in English is determined so that a heavy syllable (the head of a foot) is more prominent than any light syllable. Second, if a(n intonational) phrase contains more than one foot, any accented syllable needs to be more prominent than a heavy syllable that is not accented.

However, not all heavy syllables need to be accented. A final point is that, although every intonational phrase must contain at least one accented syllable, it can contain more than one, in which case the last accented syllable (bearing a nuclear accent) is the most prominent syllable of the phrase. Beckman & Edwards (1994) have proposed that each of these prominence types is signaled by different dominant acoustic cues. Accordingly, stressed (or heavy) syllables are distinguished from unstressed (or light) syllable by quality and duration. Nuclear accented syllables are distinguished from non-accented stressed syllables by an *F0* marker, that is, by a pitch accent. The choice of terminology above reflects language specific features, but the point is that prominence levels are assigned both on a sentence and word level depending on the prosodic phrase structure. However, there is ample evidence that languages differ from each other in the acoustic realization of prominence (see Hirst & Di Cristo, 1998, for an elaboration of this point). For example, duration and vowel quality are not reliable acoustic correlates of stressed vs. unstressed syllables in Finnish, first, because there is no linguistically determined vowel reduction in Finnish (Iivonen, 1998). Second, due to phonological length in Finnish (see above) durational differences between phonemes need to be carefully controlled in order not to change the meaning of the word. One consequence of language specificity of the relationship between prosody and its acoustic realization is that a development of a universal prosodic transcription system, similar to IPA for segmental transcription, has not been successful. Several systems for transcription of intonation patterns have been suggested such as ToBI (Silverman et al., 1992) and INTSINT (International Transcription System for Intonation; Hirst & Di Cristo, 1998), but there is no widely accepted consensus regarding the use of these systems.

In the current thesis, it will be argued that relative prominence levels between syllables are important in signaling word boundary in spoken Finnish. A hypothesis is put forth that, besides prominence related to a (nuclear) accented syllable, also the prominence of a non-accented stressed syllable is sufficient to signal word boundary in spoken utterances in Finnish. These syllables, in the majority of cases, are the first syllables of the word. In the following chapters, terms such as lexical stress, (word) stress, (sentence) accent and prominence are used in a theory neutral manner. Lexical stress is an abstract linguistic unit, which specifies which syllable of the word receives the primary stress. Usually, this syllable is also perceptually the most prominent syllable of the word. For communicative purposes, syllables may also receive sentence accent. The location of the accented syllable is usually, but not always, the lexically stressed syllable. An example of sentence accent on an unstressed syllable can be found when that syllable is contrasted as in the sentence: "I said *cofFIN*, not *cofFEE*". The accented syllable is usually the most prominent (and thus, perceptually the most salient) syllable of the utterance. The utterance

usually also has other stressed (or prominent) syllables. Thus, one critical difference in the concept of word and sentence level prominence is that there are fewer prominence levels within the word than there are within the sentence domain. In the following chapters, stress and prominence are used interchangeably; however, these concepts should be kept apart from the concept of lexical stress.

CURRENT APPROACH

Word stress and vowel harmony as cues to word boundary in Finnish language

The crucial question regarding the phonological properties of Finnish described above is: “Do these properties have functional consequences?” Are listeners actually sensitive to these cues? The second question concerns the situation in which both of these cues are available at the same time: “Are all cues equal or do listeners prefer some cues to the other?” And the third question entertained in this thesis is, “If some cues override other cues, in what condition does this happen?”

Suomi et al. (1997) reported that vowel harmony information is exploited by Finnish listeners when they segment fluent speech. Their results were expanded by Vroomen et al. (1998; see also Chapter 2 in this thesis) who showed that vowel harmony is a language specific feature, as Finns, but not Dutch and French listeners, benefited from vowel harmony information. More importantly, they also showed that the stress (or prominence) cue can override vowel harmony.

One illustration of how word stress and vowel harmony might be realized in continuous speech is depicted in Figure 2, which displays the intonation curve of the sentence “*Marja näkee ikkunan läpi äidin*”. It shows that most stressed syllables have a higher pitch (rise in the F_0). It is important to note that not all word initial syllables are marked with F_0 movement for stress. In this example the finite verb (*näkee*, ‘sees’) and the adverb (*läpi*, ‘through’) do not contain a pitch accent (see also, e.g., Välimaa-Blum, 1993). These two word boundaries that are not signaled by pitch change, contain (probably just by coincidence) a vowel harmony mismatch (that is, “*Marja näkee...*” and “*...ikkunan läpi...*”; /a/ and /æ/ belong to different harmony classes).

The present thesis describes how these kinds of multiple cues are used in speech segmentation.

A word about methodology

The answers to the questions mentioned above were searched by using a variety of different methodologies including reaction time studies, (off-line) listening tasks and recording of the brain's electric activity (event-related potentials, ERPs). Furthermore, phonetic techniques were used both to analyze the acoustic structure of the stimuli employed on some of the experiments, and also to synthesize new stimuli. The rationale behind this approach was twofold. First and foremost, if results from studies using different methodologies converge, then the arguments put forward will gain more support than just by resorting to a single method. Second, different methods have different advantages in terms of applicability, ease of use, or how accurately the cognitive processes can be followed in real time. Accordingly, all these methods should be seen as complementary.

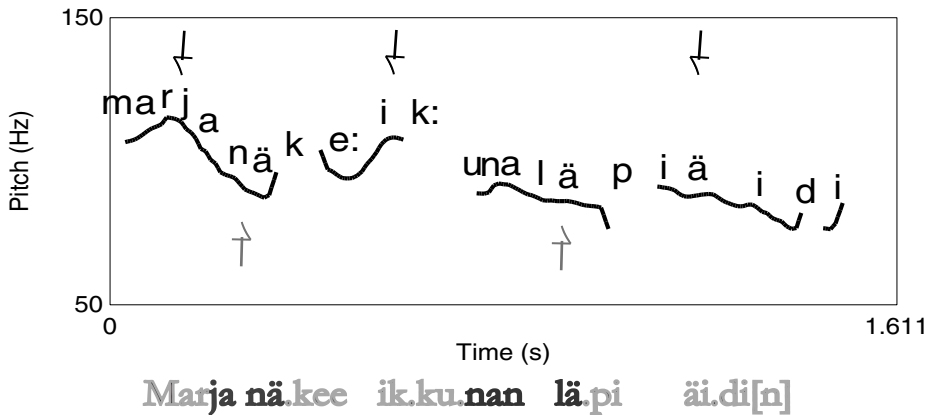


FIGURE 2. F0 contour of the utterance *Marja näkee ikkunan läpi äidin* ('Marja sees mother through the window'). Black arrows indicate rises in F0, which coincide with the first syllables of words. Note that not all word initial syllables have a pitch rise. The verb and the adverb have not received a rise in the F0 contour (gray arrows).

Behavioral measures: word spotting and artificial language learning

In the study of lexical segmentation, the reaction time speed and response accuracy of the participant have often been measured by employing a 'word spotting' task (McQueen, 1996), which is a variant of the lexical decision task. In a typical design, participants hear a list of pseudowords, which sometimes contain a real word

embedded in them. The real word can be at the onset or at the end of the pseudoword string. The participants' task is to press a response button as fast as possible if they hear a real word, and then say aloud what that word was. Word spotting differs from the word monitoring task in that the participants do not know in advance what the target words are. Instead, they try to spot any real word. In many of the previous studies using word spotting, stimuli have been bisyllabic pseudoword strings consisting of a word plus a nonsense syllable. In the current research tri-syllabic pseudoword strings were used in all experiments. Word spotting has a certain ecological validity, which relates to the fact that listeners usually recognize words in continuous speech. In word spotting, listeners are required to segment continuous input, although it might be fair to characterize the pseudoword string as a 'minimal stretch of continuous speech'. Word spotting also provides a measure of the lexical activation and competition process, and requires lexical access so that real words can be identified. This aspect contrasts word spotting with, for example, phoneme monitoring, in which target detection can be performed via a lexical look-up or via non-lexical processing. However, because the participants are required to explicitly identify the real words, word spotting is ill suited for cross-language comparison. The knowledge of the target language will vary (sometimes considerably) between participants with different language backgrounds rendering a direct comparison of the results unreasonable. For this reason, in the current thesis, another method was used to address the issue of language specificity of vowel harmony and word stress as cues to word boundaries in Finnish.

One possibility to overcome the problem related to differences in the language background is to devise an artificial language (that no one speaks or knows), and manipulate experimentally the critical cues that supposedly help in segmenting the input. This type of approach has previously been used in the study of segmentation of spoken language both with adults (Hayes & Clark, 1970; Saffran, Newport, & Aslin, 1986) and children learning a first language (Saffran, Aslin, & Newport, 1996). An artificial language learning task may be characterized as follows. An experiment consists of two phases: a learning phase and a test phase. A small set of experimental items (typically less than 10) is constructed. Experimental manipulations may involve any level of linguistic structure. In previous studies and also in our study (Vroomen et al., 1998; Chapter 2), either clustering of 'sounds' or phonemes by varying their 'statistical density', or phonological features (such as the presence or absence of vowel harmony clash or location of the stressed syllable) have been the major focus. The items are then synthesized (in the early days with a general-purpose sound generator, or currently with a text-to-speech synthesizer), and concatenated together without intervening pauses. The resulting stream of 'words' is played back to listeners, and the listeners' task is just to listen to the stream

and to try to figure out which strings of sounds might constitute a word in the artificial language. The important issue is that the size of the 'lexicon' has not been revealed to the listeners. There has been a lot of variance between the experiments in the time interval that listeners have been exposed to the stimulation. For example, Hayes & Clark (1970) synthesized a four-word list (consisting of 8 and 6 phonemes long words) which was played back in random order for 40 minutes. Saffran et al. (1996) asked their participants to listen to the sequence for 20 minutes. In our experiments (Chapter 2, Experiment 3), the listening time was restricted to 10 minutes.

After the learning phase, the learning (and segmentation) performance is tested in a separate task. Hayes & Clark (1970) constructed two sets of test items. Set 'A' consisted of the experimental stimuli and set 'B' consisted of stimuli sharing the same phonemes but either in a different order or with pauses inserted in wrong places. On each trial, a sequence of four words from each set was presented. The participants were asked to indicate which set of the stimuli, A or B, sounded similar to the stimuli presented during the listening phase. In another version of the test (e.g., Saffran et al., 1996; Vroomen et al., 1998; Chapter 2), the participants were presented with pairs of stimuli. In each pair, one of the stimuli belonged to the experimental set (which they had heard during the learning phase). The other stimulus of the pair was a foil comprising the same syllables but in an order different from the experimental stimuli. The task of the participants was to indicate which of the items they had heard during the learning phase.

Earlier research using an artificial language learning task has concentrated on whether and how the statistical properties of the input help parse the continuous sequence in to meaningful units. Results have indicated that listeners exploit statistical regularities (in this case, the higher transitional probability of within-word syllable sequence as compared to between-word sequence) when they segment continuous speech. Segmentation may take place even when participants are not explicitly paying attention to the auditory input during the learning phase (Saffran, Newport, Aslin, Tunick, et al., 1997). Finally, the artificial language learning technique has shown that statistical learning plays a role in segmenting non-linguistic input. Saffran, Johnson, Aslin, & Newport (1999) found out that both adults and children learned to segment a tone stream based on statistical information. The suggestion is that a general learning mechanism may underlie segmentation exploiting statistical properties of the speech signal.

In the experiments described in Chapter 2 (Experiment 3), it is shown that the artificial language task can be used successfully to investigate how both universal and language specific mechanisms operate in lexical segmentation.

Event-related brain potentials as a tool in cognitive neuroscience

ERPs are minute changes in the voltage fluctuation buried in the background activity of the EEG during cognitive processing. Because the evoked potential to a single stimulus recorded from the scalp is very small (typically within a range of 5-10 μV), additional means have to be employed to extract the signal from the background activity. One way to do this is by using an averaging technique. Averaging enhances the signal that is time-locked to the external stimulus (or whatever event) and simultaneously reduces the random background activity thus improving the signal-to-noise ratio.

The brain activity underlying the ERPs is probably generated by the synchronized activation of the dendritic arborization of hundreds of thousands of pyramidal cells in the cerebral cortex. The activity pattern (that is, a source-sink distribution) within a patch of the cortex can be approximated by a single equivalent dipole located in the middle of the patch and oriented vertically to it. Thus, the electric activity recorded from the scalp consists of a complex pattern of spatially and temporally overlapping dipoles. Given the theoretical problems related to localizing the sources (the so-called *inverse problem*), it is doubtful that ERPs can be effectively used in localizing the generators of cognitive processing. Instead, most cognitive neuroscientists have emphasized the extremely fine *temporal resolution* of ERPs. Consequently, the major advantage of ERPs is that they provide a time window to cognitive processing on a millisecond-by-millisecond basis. Even though this time-window is indirect (that is, the interpretation of the electric response is usually based on a functional model of a specified cognitive process), ERPs are the only brain imaging method (together with its magnetic counterpart, magnetoencephalography, MEG) that allows us to follow processing on-line as it happens in the brain. Accordingly, ERPs can be used to test hypotheses generated by cognitive theorists (see e.g., Garnsey et al. 1989). Furthermore, electrophysiological studies may yield results that force theorists to reformulate their models (King & Kutas, 1995). All in all, the ERP technique should be regarded as complementary to (more traditional) behavioral measures, and this is exactly the approach that is bolstered in the current thesis. (See Kutas and van Petten, 1994; Kutas and Dale, 1997; Kutas, Federmeier, & Sereno, 1999, for recent reviews on using ERPs in cognitive neuroscience and in particular, in psycholinguistic research on language processing).

Electrophysiology of spoken language processing

Most of the ERP research using language stimuli has centered on the comprehension of written words typically presented in sentence context. The results of several studies suggest that modality specific processing yields differences in the early

components during the first 200 – 300 ms after the onset of the stimulus, which can be seen most readily in differences in the scalp potential distribution, and also in the timing of the early components. Auditorily presented words evoke a broadly distributed N1 and a smaller P2. This complex occurs between 80 and 220 ms. N1 and P2 are considered as exogenous components, which are affected by the physical parameters, such as the (fundamental) frequency and intensity of the acoustic signal.

The following fairly slow negative deflection has been correlated with the processing of the phonological and semantic aspects of the spoken stimuli. All language stimuli evoke a negativity peaking around 400 ms (so-called N400). Most typically N400 is elicited by semantic anomalies both in written (Kutas & Hillyard, 1980) and spoken language (McCallum et al., 1984), although phonological manipulations have been suggested to affect the amplitude of N400 (e.g., Rugg, 1984; Praamstra et al. 1994). N400 to spoken words is more sustained over frontal than posterior sites (Holcomb & Neville, 1990), and may be larger over the left than the right hemisphere electrodes (Kutas & van Petten, 1994). The similar overall appearance of the auditory and visual N400 as well as the fact that N400 is also evoked by semantic incongruities in (American) Sign Language (Neville, Mills, & Lawson, 1992) have been taken as an index of the workings of an amodal semantic system (Holcomb & Neville, 1990; Holcomb & Anderson, 1993). In most reports N400 has been attributed to controlled post-lexical integrative processes as opposed to an automatic process of (spreading activation in) lexical access. In this respect, one of the most compelling findings regarding the functional locus of N400 was reported by Brown and Hagoort (1993) using a masked priming paradigm. They showed that the N400 could be recorded only when the prime was not masked. In contrast, a significant behavioral priming effect was obtained both in the unmasked and masked conditions. This suggests that N400 is sensitive to post-lexical integration, and does not reflect automatic spreading of lexical activation.

The N400 complex also shows modality specific patterns, of which the earlier onset and longer lasting negativity are typical of auditory modality. The early part of the auditory N400 was manifested in some reports as a clear and distinct negativity peaking around 200-250 ms post stimulus onset. Some researchers have related the early part to phonological processing, and results suggest that phonological and semantic effects are dissociable (Connolly and Phillips, 1994). However, there is evidence that semantic processing may have an effect already around 200 ms post stimulus onset when words are presented in sentence context (Van Petten et al., 1999). The earlier onset of auditory N400 as compared to the visual counterpart informs, first, about the temporal nature of spoken language. That is, sounds that constitute the word are exposed to the listener (almost) serially in a left-to-right

manner over a period of several hundreds of milliseconds. In addition, the pattern tells us that spoken word recognition starts before the acoustic offset of the word (Marslen-Wilson, 1973; Holcomb & Neville, 1990; van Petten et al., 1999). Finally, Hagoort and Brown (2000) have suggested that the early negativity observed to spoken words reflects two related processes. First, surplus negativity is evoked by a (phonological) mismatch between the expected word form on the basis of the context (see also Connolly & Phillips, 1994), and second, the extra negativity indexes the activation (and possibly the competition) of the lexical candidates that are generated by the acoustic input.

The latter part of the ERP waveform consists of the late positive component (LPC), or Slow Wave (SW) that is usually present both when written and spoken language stimuli are used. Functionally, the SW is considered as a member of the so-called P300 family reflecting post-lexical and controlled processing related to expectancy, attention, decision making, or context updating (Coulson et al., 1998; Donchin & Coles, 1988). Some researchers have regarded the SW as an index of the processing load demands induced by the task. For example, in a priming study Brown, Hagoort, and Chwilla (2000) suggested that, irrespective of the task demands, participants always try to construct an integrated representation of the word pair. Thus, finding a link between a prime and the related target is readily available. Consequently, the processing load is low and yields a more positive SW. For unrelated and neutral pairs, the link between the prime and target is harder to obtain, which is shown in a more negative SW. One should note that the SW is present most readily when an additional task (such as pressing the response button or counting specific targets etc.) is required. If participants are engaged in a more natural task such as listening for understanding, the slow wave is practically non-existent (for example, compare Figs. 3 and 6 in Brown et al., 2000). Finally, in paradigms in which morpho-syntactic aspects of the stimulus are manipulated, some researchers have correlated the SW (or P600) with second-pass parsing (that is, reanalysis and repair) of garden-path sentences (see, for example, Hahne & Friederici, 1999; see Coulson et al., 1998, for a different view). However, all different interpretations explicitly or implicitly refer to post-lexical processing as underlying the late positive wave.

To my knowledge, there are no reports directly focusing on lexical segmentation of spoken language using ERPs. A recent report by Böcker et al. (1999) investigated the ERP correlates of metrical stress in Dutch. As already mentioned earlier, metrical stress has been shown to facilitate lexical segmentation by providing a cue about the word boundary. In the experiment by Böcker et al., bisyllabic words presented in isolation with different metrical structure (Weak-Strong (WS) or Strong-Weak (SW)) were used as stimuli. Thus, no segmentation was required by

the participants. However, differences between conditions as a function of the metrical structure were observed. Specifically, WS words evoked a more pronounced negativity around 300 ms postonset of the stimulus as compared to SW words. They interpreted this negativity as an index of extra processing required by a less typical prosodic pattern (WS) in Dutch. In another study (Böcker et al., submitted) the same ERP pattern was also observed with pseudowords. In a condition in which the experimental real word stimuli were low-pass filtered the negativity was delayed by approximately 100 ms. Böcker et al. concluded that the main acoustic correlate of metrical stress is vowel color and not the other acoustic parameters such as intensity or fundamental frequency (see also Fear, Cutler, & Butterfield (1995) for a similar account of metrical stress in English).

ORGANIZATION OF THE THESIS

Chapter 2 introduces the basic findings. First, both word stress and vowel harmony facilitate lexical segmentation in Finnish. Second, it appears that word stress is the primary cue and vowel harmony (mismatch) is a secondary one. If word stress provides correct information about the word boundary, it is the main cue that listeners use to detect words in continuous speech. If, however, word stress is missing or if it provides conflicting information about the word boundary listeners rely on vowel harmony.

In Chapter 3, the acoustic correlates of word stress are explored. The impetus behind this study relates to the fact that, in Finnish, lexical stress is fixed, which may yield it redundant, and it may not be realized acoustically. The acoustic analysis of the experimental stimuli used by Vroomen et al. (1998; Chapter 2 in the present thesis) revealed that the fundamental frequency (F_0) seems to be the most important acoustic correlate of prominence in Finnish at least in the material used in that study. The suggestion is that F_0 is one component underlying the rhythm that listeners exploit in detecting the onsets of words. The exact nature of the pitch movement was left open, as different types of variables derived from the F_0 analyses were equally successful in predicting the behavioral performance.

In Chapter 4, the research question is whether sentence accent is needed to signal prominence, or does word stress suffice. This relates to the previous discussions on the acoustic realization of word stress in Finnish. It could have been that in the materials used by Vroomen et al. (1998) the prominence of syllables was realized by a larger pitch movement, typical of sentence accent. It may be that word stress has usually a smaller movement. Since acoustic details on this issue are nonexistent, this hypothesis was tested by asking the participants in a word spotting task to detect

target words that were excised from an unaccented position in a sentence; thus target words lacked sentence accent. Results showed that detection speed and accuracy were comparable to those obtained by Vroomen et al. (1998) in their Experiment 2, in which the first syllable of the target word was the most prominent. These results strongly suggest that sentence accent is not needed, and that word stress is sufficient to cue a word boundary.

Chapter 5 focuses on the temporal aspects of lexical segmentation. To this end, both reaction times and ERPs were measured. More specifically, we wanted to explore the time course of the processing of the two potential cues to word boundaries. The reaction time results of the current study replicated the earlier results of Suomi et al. (1997) and Vroomen et al. (1998) in which real word targets were used. Our new findings were related to the pseudoword targets, which showed both facilitatory effects of stress position (faster responses to stressed targets as compared to unstressed targets) and vowel harmony (faster responses to disharmonious as compared to harmonious targets). However, the pseudowords profited from correct stress information more than the real words, suggesting that word stress is computed pre-lexically. The ERP results confirmed the behavioral ones. In addition, the ERPs indicated that the detection of vowel harmony mismatch began before the acoustic offset of the target item and continued several hundred milliseconds post stimulus offset, possibly suggesting that post-lexical processing was present. Differences in the vowel harmony effect between the real versus pseudoword targets were obtained. The distribution of the negative scalp potential to the real word targets was much larger involving the left hemisphere while for the pseudowords only weak effects were observed suggesting that different (cortical) areas were responsible for this effect.

In Chapter 6, a summary of the results is presented. Some suggestions are also given as to how to model lexical segmentation using multiple cues. The focus will be on how lexical stress and vowel harmony in Finnish might be incorporated into a model of spoken word recognition. At the end of Chapter 6 a few suggestions for future research and for possible (clinical) applications are provided.

CHAPTER 2

THE ROLES OF WORD STRESS AND VOWEL HARMONY IN SPEECH SEGMENTATION¹

ABSTRACT

Three experiments investigated the role of word stress and vowel harmony in speech segmentation. Finnish has fixed word stress on the initial syllable, and vowels from a front or back harmony set cannot co-occur within a word. In Experiment 1, we replicated the results of Suomi, McQueen, and Cutler (1997) showing that Finns use a mismatch in vowel harmony as word boundary cue when the target-initial syllable is unstressed. Listeners found it easier to detect words such as *HYmy* in *PUhymy* (harmony mismatch) than in *PYhymy* (no harmony mismatch). In Experiment 2, words had stressed target-initial syllables (*HYmy* as in *pyHYmy* or *puHYmy*). Reaction times were now faster and the vowel harmony effect was greatly reduced. In Experiment 3, Finnish, Dutch, and French listeners learned to segment an artificial language. Performance was best when the phonological properties of the artificial language matched those of the native one. Finns profited, as in the previous experiments, from vowel harmony and word-initial stress; Dutch profited from word-initial stress, and French neither profited from vowel-harmony nor from word-initial stress. Vowel disharmony and word-initial stress are thus language-specific cues to word boundaries.

INTRODUCTION

One of the major issues in spoken word recognition concerns the detection of word boundaries in continuous speech. The central problem is to understand how listeners segment the continuous speech signal into discrete words when there are no reliable acoustic cues that signal the beginnings of words. A number of alternative ideas have appeared in the literature that point toward a possible solution. A major

¹ This chapter has been published in *Journal of Memory and Language*, 38, 133-149 (Vroomen, J., Tuomainen, J., and de Gelder, B. (1998).

division can be made between proposals that emphasize acoustic/phonetic cues and those that focus on lexical or contextual processes. In the former, word boundaries are located on the basis of local perceptual features such as the presence of glottal stops, laryngealized voicing, increased aspiration, or vowel lengthening (e.g., Lehiste, 1960; Nakatani & Schaffer, 1978). Proposals in the latter category use concepts such as the uniqueness point of the word, lexical competition, or 'top-down' knowledge (e.g., Cole & Jakimik, 1980; Marslen-Wilson, 1984; McClelland & Elman, 1986; Norris, 1994).

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

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

In the present study, we explored some of these issues by examining word stress and vowel harmony as potential segmentation cues in Finnish. Finnish has front-back vowel harmony (Karlsson, 1983). The Finnish vowels /u, ɑ, o/ belong to the back harmony set, /y, æ, ø/ to the front harmony set, and /i, e/ are neutral. The main restriction in uncompounded Finnish words is that vowels from the front and the back harmony class cannot occur together, but vowels from the neutral class can be combined with both the front or back class vowels in any position in the word stem.

Harmony propagates from left to right from the first vowel in the root to subsequent vowels in root and suffix. Vowels of suffixes are therefore subject to the harmony restriction. As an example, *kapula* (meaning *stick*) and *räjähdyks* (*explosion*) are possible Finnish words because /a, u/ are from the back harmony class, and /æ, y/ are from the front harmony class (/æ/ is written as *ä*). The correctly suffixed forms of the words would be *kapulako* (meaning *a stick?*) and *räjähdyksö* (*an explosion?*, /ø/ is written as *ö*). But *kapyla* and *räjähdys* would be prohibited as Finnish words because their vowels are from opposing classes. A clash in vowel harmony (for example a front vowel followed by a back vowel, or vice versa) is in Finnish thus typically associated with a word boundary. (There are some exceptions to this rule such as *analyysi*, meaning *analysis*.)

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

The potential use of stress as a cue to word boundaries was studied recently by Iivonen, Niemi, and Paananen (submitted), who tried to determine the extent to which fundamental frequency (F0) peaks in Finnish, English, and German coincide with word stress. They analyzed TV and radio newscasts and counted how often a just noticeable F0 peak (defined as a difference in one semitone or more when compared with the neighboring syllable) matched a primary stressed syllable. One cannot expect a perfect correlation between F0 peaks and word stress because stress may not always be acoustically realized. In addition, not every F0 peak signals word stress, because it is well known that the F0 contour has other linguistic functions such as accentuation, signaling emotions, or cueing syntactic boundaries (see Cutler, Dahan, & van Donselaar, 1997 for a recent overview). These and other rhythmic

phenomena such as the avoidance of stress clashes are likely to obscure the relation between word stress and its phonetic correlates. Nevertheless, Iivonen, Niemi, and Paananen found that the majority of Finnish F_0 peaks, 73%, occurred on the primary stressed syllable, while only 42% of the German peaks and 59% of the English peaks represented word stress. Moreover, about 52% of the Finnish word-initial syllables had an F_0 peak. Thus, this phonetic analysis suggests that F_0 peaks are at least partly successful in signaling where primary stress is, and hence, where a word boundary is located in Finnish speech.

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

In our view, the role of word stress in speech segmentation is still a matter of debate because so far little is known about the role of word stress in different languages. Moreover, the presumption that word stress can only be determined post-lexically may be wrong. It seems possible that a stressed syllable can be perceived as stressed without reference to neighboring syllables, for example on the basis of characteristic F_0 transitions within the syllable, a long duration, or an increased intensity (of higher harmonics). In addition, a stressed syllable in continuous speech may stand out relative to the previous syllable. Given that almost all Finnish words are multisyllabic with unstressed final syllables, stressed syllables are usually preceded by the unstressed word-final syllable of the preceding word. For these reasons, stressed syllables may be perceived as stressed even though the word to which they belong is not yet recognized. There is therefore no strong a priori reason to rule out word stress as a segmentation cue.

To investigate the combined roles of word stress and vowel harmony in speech segmentation, we conducted a study in which both factors were varied. Experiment 1 was a replication of Suomi et al. (1997, Experiments 1 and 4) in which word boundaries did not have a stress cue. Listeners had to detect words such as *HYmy* (the stressed syllable is denoted with capital letters) in *PUhmy* (harmony clash

between prefix and target word; no stress cue on the first syllable of the embedded word) or *PYhymy* (no harmony clash; no stress). This replication was conducted first in order to have a basis for later comparisons. It also allowed us to check whether we had artifacts in items, subjects, equipment, or procedures that might explain any deviant results. Experiment 2 was similar to the previous one, except that target words now contained a stress cue such as *HYmy* in *puHYmy* or *pyHYmy*. In Experiment 3, we used an artificial learning task in which Finnish, French, and Dutch speakers had to segment an artificial language into ‘words’. This allowed us to test the generality of our findings across different tasks and to examine the extent to which vowel harmony and word stress are language-specific cues to word boundaries.

EXPERIMENT 1

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

METHOD

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

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

which contained any other word besides the intended one. The target-bearing items are listed in the Appendix.

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

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

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

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

RESULTS AND DISCUSSION

Unless stated otherwise, all analyses were done in exactly the same way as by Suomi et al. (1997). Reaction times (RT) were measured from the offset of the word, and vocal responses that did not correspond to the intended word (0%) and outlying responses (4%) were treated as errors and discarded from the RT analyses. Outlying responses were defined as RTs slower than 2000 ms as measured from target offset. It should be noted that Suomi et al. used the same upper cut-off criterion, but they also discarded RTs faster than 150 ms. In our Experiment 1, no response was faster than this criterion. However, in our Experiment 2 responses were much faster, and in that case it would not have been correct to treat RTs faster than 150 ms as 'outliers'. For consistency across our experiments, we therefore discarded only responses longer than 2000 ms. The false alarm rate (i.e., a key

response on a filler item) was 2.1%. Inspection of individual items and subjects showed that no item was missed by more than 50% of the subjects and no subject made more than 50% errors. No subject or item was therefore excluded. The mean RTs and error rates are presented in the top panel of Table 1.

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

TABLE 1. Mean Reaction Time (in ms) and Error Rate (in parentheses) in Experiment 1 and Suomi et al. (1997)

EXPERIMENT 1				
CONTEXT	RT FROM TARGET OFFSET Target		RT FROM TARGET ONSET Target	
	BACK	FRONT	BACK	FRONT
HARMONIOUS	870 (12%)	891 (15%)	1228 (16%)	1206 (22%)
DISHARMONIOUS	779 (9%)	688 (9%)	1122 (13%)	1042 (10%)
SUOMI ET AL.				
HARMONIOUS	802 (9%)	822 (10%)		
DISHARMONIOUS	699 (5%)	604 (4%)		

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

Analysis of the error rates showed no trend of a speed-accuracy trade-off. The ANOVA on the errors by subjects showed that more targets were missed when the prefix was harmonious than when it was disharmonious (13% vs. 9%), $F(1,19) = 4.93$, $p < .05$, but this difference was not significant in the item analysis, $F(2) < 1$. No other main effect or interaction was significant (all *F*'s < 1). This error pattern is again very similar to that of Suomi et al. (1997). In their Experiment 1, they found a significant main effect of prefix in the same direction as ours, but no other effects were significant.

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

As a further control for the duration of the items, we measured RTs from word onset (see Table 1). In this analysis, we again discarded RTs longer than 2000 ms, this time measured from word onset. This follows Suomi et al. (1997), even though it is debatable whether the same cut-off criterion of 2000 ms can be justified because more RTs than in the previous analyses had to be discarded (8% versus 4%). There was a harmony effect of 135 ms which was significant by subjects only, $F(1,19) = 59.28$, $p < .001$, $F(1,28) = 2.32$, $p = .13$. The interaction with harmony class of the target was not significant, $F(1,19) = 1.11$, ns; $F(2) < 1$. Pairwise comparison showed that the harmony effect was significant in the subject analysis for targets with back vowels, $F(1,19) = 7.97$, $p < .02$, and for targets with front vowels, $F(1,19) = 36.05$, $p < .001$, but the effects were not significant in the item analysis (both p 's $> .10$). Thus, the results of the item analyses in which RT was measured from word onset were somewhat weaker than those in which RT was measured from word offset, but this is understandable because more RTs were discarded that passed the time-out criterion. The results are again similar to the results of Experiment 1 of Suomi et al. (1997) in which there was also no significant interaction in the item analysis. No comparison can be made with their Experiment 4, because these analyses were not reported.

We also performed a new analysis on the error rates because more responses passed the 2000 ms time-out criterion. The subject analysis now showed that more errors were made with a harmonious prefix (19%) than with a disharmonious prefix (11%), $F(1,19) = 8.72$, $p < .001$, but this difference did not reach significance in the item analysis, $F(1,28) = 1.60$, NS). There was also a significant interaction in the subject analysis between prefix type and harmony class of the target, $F(1,19) = 4.38$, $p < .05$; $F(2) < 1$, showing that the difference between a harmonious and a disharmonious prefix was bigger in targets with vowels from the front harmony class (12% difference) than in targets with vowels from the back harmony class (3%).

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

EXPERIMENT 2

In Experiment 2 we investigated whether word stress plays a role in speech segmentation and whether the vowel harmony effect remains the same when the

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

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

METHOD

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

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

RESULTS AND DISCUSSION

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

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

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

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

The crucial analysis is the comparison between Experiment 1 and 2, because that will show whether stress had an effect and whether it changed the harmony effect. An ANOVA was conducted on the RTs in which Experiment was a between-subjects and a within-items factor. When RTs were measured from word offset,

there was a main effect of Experiment because RTs were much faster in Experiment 2 than in Experiment 1, $F1(1,38) = 66.38$, $p < .001$; $F2(1,28) = 984.56$, $p < .001$. There was also an interaction between Experiment and harmonious/disharmonious prefix showing that the harmony effect was present in Experiment 1 (147 ms), but not in Experiment 2 (0 ms), $F1(1,32) = 31.57$, $p < .001$; $F2(1,28) = 4.14$, $p = .05$. When RTs were measured from word onset, there was again a main effect of Experiment, $F1(1,38) = 59.64$, $p < .001$; $F2(1,28) = 731.13$, $p < .001$. The interaction between Experiment and the harmony effect was significant by subjects only, $F1(1,38) = 41.37$, $p < .001$; $F2(2,28) = 2.28$, $p = .14$.

TABLE 2. Mean Reaction Time (in ms) and Error Rate (in Parentheses) in Experiment 2.

Context	RT FROM TARGET OFFSET		RT FROM TARGET ONSET	
	<i>Target</i>		<i>Target</i>	
	BACK	FRONT	BACK	FRONT
HARMONIOUS	270 (5%)	285 (9%)	696 (5%)	712 (9%)
DISHARMONIOUS	286 (5%)	270 (9%)	702 (5%)	678 (9%)

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

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

EXPERIMENT 3

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

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

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

To determine whether listeners indeed apply the native segmentation routine when performing the learning task, we presented the same materials to listeners from different language backgrounds. For the present comparison, French is maximally different from Finnish because French does not have vowel harmony, and stress in French polysyllabic words is never on the initial syllable but always on the last full vowel of content words (Dell & Vergnaud, 1984). If the task reflects properties of

the native segmentation routine, then French listeners should not be influenced by whether words are harmonious or disharmonious. Also, word-initial stress should not be helpful because that conflicts with the French stress pattern.

An intermediate case between Finnish and French is Dutch. Dutch, like French, has no vowel harmony. We therefore expected Dutch listeners not to be sensitive to vowel harmony. The position of the stressed syllable is, unlike Finnish and French, variable in Dutch. According to Kager (1989), the penultimate position receives primary stress as default, but a count in the Dutch CELEX lexicon showed that most multi-syllabic words have stress on the initial syllable. Of all two-, three-, and four-syllabic words with a frequency of occurrence higher than or equal to one, 56% of the tokens had lexical stress on the first syllable (15,357 entries out of 27,020 selected words). For tri-syllabic words as were used in the present experiment, 53% percent had stress on the initial syllable (6,220 words out of all 11,646 trisyllabic words), 32% (or 3,788 words) had stress on the penultimate syllable, and 14% (1,638 words) had stress on the final syllable. Taking these statistical facts into account, stressed syllables are likely to be a word onset in Dutch, and Dutch listeners may therefore profit from a stress cue on the word-initial syllable.

METHOD

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

Materials. For the learning phase, an artificial language was constructed consisting of four consonants (/v/, /m/, /t/, and /k/) and six vowels (/o/, /u/, /a/, /y/, /ɛ/ and /æ/) that made up 15 different CV syllables. The syllables were combined so as to create two separate lexicons, a harmonious and a disharmonious one, each consisting of six trisyllabic words. The words in the harmonious lexicon had vowels belonging either to the front harmony set (/y/, /æ/ and /ɛ/) or the back harmony set (/u/, /o/ and /a/). The back harmony words were /vomuvu/, /tokuvo/, /motamu/; the front harmony words were /mymety/, /vykeve/, /tykety/. The words in the disharmonious lexicon were created by replacing one or two vowels of the harmonious words so that /o/ became /æ/, /ɛ/ became /a/ and /u/ became /y/. This resulted in the words /væmyvu/,

/tokuvœ/, /motamy/, /mumety/, /vykavε/, /tykaty/. None of the items was, in any obvious sense, similar to a real Finnish, French, or Dutch word.

For both lexicons (harmonious and disharmonious) two versions with a different stress pattern were created. In the no-stress versions, all the words' syllables had equal stress, whereas in the stress-initial versions, the first syllable of each word received a pitch accent. This resulted in four experimental versions. Each version consisted of 150 tokens of the six words (total of 900 words, 2700 syllables). The words were concatenated in random order without spaces into a text file with the restriction that the same word could not occur twice in a row. The four versions had the same random order. The text file was split into 5 blocks of equal length, and each file was then input to the Spengi text-to-speech synthesizer at the Institute for Perception Research (IPO) in Eindhoven, which is based on Dutch diphone synthesis. The synthesizer speech rate was adjusted to a natural speech rate of approximately 275 syllables per minute. The phoneme durations were kept constant in all versions. In the no-stress version, the fundamental frequency was kept monotonous at 120 Hz throughout the whole string. In the stress-initial version, stress on the initial syllable was acoustically realized by using a pitch accent. The synthesis parameters for the F_0 were set to its default values. The F_0 linearly increased on the first syllable from 120 Hz to 170 Hz, and then gradually decreased over the next two syllables back to baseline². The synthesizer output was saved on an audio file (AIFF format, 16 bit precision, 16 kHz sampling rate), and each file was then recorded directly from a Silicon Graphics Iris Indigo workstation on a DAT tape.

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

² Saffran, Newport and Aslin (1996) used lengthening of the vowel as a cue for lexical stress. With English, they did not find an improvement when the word-initial vowel was lengthened. We conjecture that, at least for Finnish and Dutch, pitch accent is a better realization of stress for word-initial syllables than vowel lengthening (see for example 't Hart, Collier, & Cohen, 1990.)

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

Design and Procedure. Participants were tested in groups of 2 to 8. As far as possible, equal number of listeners received one of the four versions. They were instructed to listen to the nonsense language and were told that the language consists of 'words' with no meaning or syntax. Their task was to figure out what the words were. They were given no information about the length or the number of words. During the learning phase, they were asked to listen to five blocks of two minutes each. There was a 5 sec pause between the blocks. Participants were told that at the end a word recognition test was to be administered. The test was a two-alternative forced-choice task. Each test trial started with a tone, followed by a pair of trisyllabic strings separated by 500 ms of silence. One of the strings was a word of the artificial language, the other was one of the foils. Participants were asked to indicate whether the word came in first or second position by circling a "1" or "2" on a prepared answer sheet. They were told to guess if unsure and they were given 4 s for this. The complete test consisted of 36 trials (six words exhaustively paired with the six foils) with a short break in the middle. Four practice trials were given to acquaint participants to the structure of the test.

RESULTS

The percentage of correctly recognized words in the two-alternative forced-choice test was computed for each listener. Table 4 presents the means across subjects. Simple *t* test showed that performance in each of the twelve cells was significantly above chance (all *p*'s < .05 with a chance level of .5). An overall ANOVA with native language, stress, and vowel harmony as between-subjects factors showed that there was a main effect of language, $F(2,134) = 14.87$, $p < .001$, a main effect of stress, $F(1,134) = 20.65$, $p < .001$, and a significant interaction between language and stress, $F(2,134) = 3.33$, $p < .05$. The effect of vowel harmony and all other interactions with vowel harmony were not significant. Separate ANOVAs for each language group showed that Finns, $F(1,39) = 19.86$, $p < .001$, and Dutch, $F(1,49) = 10.83$, $p < .002$, profited from stress, but the French did not ($F < 1$). Inspection of Table 3 shows that in the Finnish group there was a trend toward an interaction between stress and vowel harmony in the predicted direction, but this trend was statistically not significant, $F(1,39) = 1.11$, $p = .299$. Despite the lack of a significant interaction, separate *t*-tests were conducted because the between-subject design is statistically rather conservative. However, *t*-tests in which the harmony effect is

tested should be interpreted with caution, because the harmony effect or its interaction was not significant in the overall ANOVA.

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

VOWEL HARMONY	FINNISH		DUTCH		FRENCH	
	NO STRESS	STRESS	NO STRESS	STRESS	NO STRESS	STRESS
HARMONIOUS	73%	86%	65%	79%	58%	58%
DISHARMONIOUS	64%	85%	64%	75%	62%	67%

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

In the stress-initial condition, there was no difference between harmonious and disharmonious items, $t(17) = -.13$, NS. With harmonious items, 8 out of 9 participants (89%) performed better than chance, and with disharmonious items 9 out of 10 participants (90%), $\chi^2_{(1)} < 1$. Moreover, average performance in the stress-initial conditions was much better than in the no-stress conditions. Overall performance increased from 69% in the no-stress conditions to 86% in the stress-initial conditions; an increase of 16%. Simple t -tests showed that the improvement was significant for harmonious, $t(20) = 2.47$, $p < .02$ and disharmonious items, $t(19) = 3.80$, $p < .001$.

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

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

BETWEEN-LANGUAGE COMPARISONS

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

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

Dutch versus French. There was no difference between Dutch and French with harmonious and disharmonious no-stress items (all p 's $> .10$). However, stress-initial

harmonious items were recognized better by the Dutch than by the French, $t(21) = 3.58$, $p < .002$, $\chi^2_{(1)} = 5.06$, $p < .05$. The better performance of the Dutch with stress-initial disharmonious items failed to reach statistical significance, $t(23) = 1.21$, $p = .24$, $\chi^2_{(1)} = 2.93$, $p < .10$.

DISCUSSION

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

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

GENERAL DISCUSSION

In three experiments, we observed that Finns use, in an interdependent way, vowel disharmony and word stress as cues to word boundaries. In a word spotting task, vowel disharmony was used when the word-initial syllable was unstressed, but the effect was greatly reduced when there was a stress cue on the word-initial syllable. The same pattern was obtained in a learning task: Finns found harmonious words without a stress cue easier to segment than comparable disharmonious words, but the presence of a stress cue improved performance and the difference between harmonious and disharmonious words disappeared. These results are in direct contrast with the conclusion of Suomi et al. (1997), who argued that ‘word stress may not play an important role in recognition of Finnish speech’. They further

stated that ‘It is very unlikely that the harmony mismatch effects emerged because of the absence of canonical stress cues’. It now seems clear that this conclusion cannot be maintained. In fact, the opposite is the case: Stress is the strongest cue, and it greatly reduces the effect of vowel harmony. The results of Suomi et al. can therefore not be generalized to normal fluent speech where stressed syllables are often signaled by F_0 peaks or other stress cues (see Iivonen et al. submitted).

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

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

a priori success rate of a vowel disharmony algorithm is much lower than that of a stress based algorithm.

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

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

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

language-specific segmentation routine to be tuned. It therefore seems that an off-line task can be informative about on-line processing.

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

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

CHAPTER 3

FUNDAMENTAL FREQUENCY IS AN IMPORTANT ACOUSTIC CUE TO WORD BOUNDARIES IN SPOKEN FINNISH¹

ABSTRACT

Fixed word stress on the first syllable may be used as a cue to word boundaries in spoken Finnish. Recently, Vroomen et al. (1998) showed that correct position of word stress facilitates word recognition in a word spotting task. We studied the acoustic correlates of stress by analyzing acoustically the stimuli employed by Vroomen et al. Regression analyses revealed that fundamental frequency (F_0) was the only acoustic variable that predicted significantly reaction times to word targets. We suggest that F_0 is an important acoustic correlate of word stress, and listeners employ prominence cue to detect word boundaries in spoken Finnish.

INTRODUCTION

Lexical stress in a language with fixed stress location displays a dilemma. Since stress is fixed, as in Finnish on the first syllable, it might provide a reliable cue to word boundaries. However, because it is fixed, its overt marking may be totally redundant and, as a consequence, it might not be acoustically realized.

Recently, both issues have received some attention. Vroomen et al. (1998) investigated the possible role of word stress in lexical segmentation of Finnish. The results indicated that when word stress was located on the correct syllable (that is, on the first syllable of the target word), subjects were much faster and more accurate in detecting target words in a word spotting task. Furthermore, no other cues to

¹ This chapter is published as Tuomainen, J., Werner, S., Vroomen, J., & de Gelder, B. (1999) Fundamental frequency is an important acoustic cue to word boundaries in spoken Finnish. In Ohala, J.J., Y. Hasegawa, M. Ohala, D. Granville, & A. C. Bailey (eds.) *The Proceedings of the 14th Congress of Phonetic Sciences*, August 1-7, 1999, San Francisco, USA, 921-923.

word boundaries (such as vowel harmony mismatch) were employed. When stress cue provided conflicting information about the possible word boundary, subjects' response speed and error rates increased dramatically, and responses were facilitated by vowel harmony mismatch. They concluded that word stress was the primary cue to word boundaries in spoken Finnish.

One acoustic correlate of lexical stress, fundamental frequency (F0), was studied by Iivonen et al. (1998) who measured F0 peaks in Finnish, German English radio and TV broadcasts. They set an arbitrary 1 semitone criterion for a perceptually relevant F0 peak that was considered above the just noticeable difference of pitch perception. The results showed that there were fewer F0 peaks in Finnish than in German or English. However, 73% of all F0 peaks in Finnish occurred on the lexically stressed syllables (that is, on the first syllable of the word). The corresponding figures in German and English were 42% and 59%, respectively. Finally, 52% of all lexically stressed syllables as opposed to 27% in German, and 40% in English contained an F0 peak. In other words, about half of the first syllables of the words in Finnish contained an F0 peak. To put it the other way, about half of the first syllables were not marked. In addition, about one fourth of F0 peaks occurred outside first syllables.

As already briefly indicated, word stress in Finnish is problematic in terms of what processing consequences there might be for detection of word boundaries. Lexical stress lands invariably on the first syllable but other syllables than the first may, for contrastive purposes, receive (sentence) accent (and as a consequence, be perceived as more prominent than the lexically stressed first syllable). Separating stress and accent (conceptually) is relevant as they are different realizations of prominence, and they may have different acoustic realizations. One way of conceptualizing stress and accent is to consider them distinct but related dimensions (e.g., Sluijter & van Heuven, 1996). Stress is determined by language *system*, which specifies the strongest syllable in a word. Accent is determined by language *behavior* and is used for communicative purposes. A stressed syllable is a potential landing site for accent placement but this may not always be so (for example, focus may change prominence from a stressed syllable to unstressed syllable, "I said SUGgest, not DIgest"). In some systems, a typical F0 movement is explicitly linked with sentence accent (Hermes & Rump, 1994). We remain neutral as to how this distinction applies to Finnish (as relevant research is lacking). In this paper we have adopted the more neutral term *prominence* (that is used here interchangeably with word stress); prominence of a syllable refers to the property of a syllable to be perceptually more salient than the neighboring syllables.

Four main acoustic parameters have been correlated with prominence (and usually also with word stress). Typically, fundamental frequency has been considered the main acoustic correlate, and a smaller role has been attributed to duration, (overall) intensity, and vowel quality (Fry, 1958). Vowel quality refers to reduction of vowel quality in unstressed syllables. Since in Finnish, there is no significant reduction, we will not consider vowel quality in this paper. In addition, one interesting suggestion in terms of the role of intensity in the perception of prominence has been offered by Sluijter and her coworkers (1996, 1997) who argued that correlating overall intensity to prominence is misleading. Increased phonatory effort during a stressed syllable may not present itself as increase in overall gain but instead as a change in the spectral tilt (or emphasis) of the amplitudes of the harmonic overtones in different frequency bands. In Dutch, this emphasis in the realization of pitch accent seems to be related most notably to the (middle) frequency band between 1000 and 2000 Hz (Sluijter & van Heuven, 1996).

In this paper, we present results from acoustic analyses of the test stimuli used by Vroomen et al. (1998). In addition, we regressed the reaction times to the measured acoustic variables. Since the main independent variable in Vroomen et al. that had an enormous effect on the reaction time latencies was the position of the stressed syllable, we assumed that the results of the regression analyses might provide information about what acoustic variable(s) underlie perception of prominence in Finnish during recognition of words in connected speech.

METHODS

Stimulus material. The stimuli were 120 tri-syllabic CVCVCV items with a CVCV Finnish noun embedded in the end of the string used in two word spotting task (for further details, see Vroomen et al., 1998). The stimuli were recorded by one of the authors (JT) in a sound treated booth on a DAT tape. The same item was spoken either with the stress landing on the first (e.g. /'ku,palo/, "palo" meaning 'fire') or on the second syllable (/,ku'palo/). All stimuli were digitized with 16 bit precision (sampling rate 22050 Hz) and saved on a hard disk for further analyses.

Acoustic analyses. F0, intensity, duration, and spectral tilt of the vowels of the first and second syllable were measured automatically by the Speech Filing System software (Vs3.3), (available on the Internet at <http://www.phon.ucl.ac.uk/resource/sfs.html>) using the following procedure. First, the onsets of all individual phonemes in each stimulus were marked and labeled, and the marks and corresponding labels were saved on an ASCII file. (Ten per cent of the total stimulus material was reanalyzed by another phonetically trained researcher to

investigate the reliability of the marking procedure. Kendall's coefficient of concordance indicated extremely good agreement between the two researchers [$W(119)=.999$, $p < .0001$].

Second, the duration of each segment was computed based on the segment marks, and the intensity and the F0 of the vowel of the first and the second syllable was measured from two points, 30 ms and 60 ms from the onset of the vowel. The mean value of these two measures was then computed and later used in the statistical analyses. Finally, as changes in the relative amplitudes of the harmonics of the vowel spectrum have been indicated to play a role in the perception of prominence (Sluijter et al. 1997), the spectral tilt was measured from four adjacent frequency bandwidths (0-.5, .5-1, 1-2, and 2-4 kHz). To account for differences in overall intensity between vowels, the dB values were transformed into ratio variables by dividing the base band (0-.5 kHz) intensity with the dB value of adjacent individual frequency bands. This resulted in three spectral tilt variables for each vowel in the first and the second syllable.

RESULTS

Comparisons of the mean values of the acoustic variables between stressed and unstressed syllables revealed extremely significant differences (all p 's $< .001$) in all variables (see Table 1 for mean values of F0, intensity and duration of the test stimuli). This was as expected since usually even small acoustic differences lead to statistically significant differences because of highly correlated patterns. Most notable changes occurred in the F0 and duration. When the F0 was compared between syllables in the same position (e.g., between stressed and unstressed first syllables, KUpalo vs. kuPAla) the difference was 8 Hz (1.4 semitones, st) in the first, and 32 Hz (5.3 st) in the second syllable. When comparison was performed within a word, the difference between the first and the second syllable was 22 Hz (3.8 st) for “KUpalo“ and 18 Hz (2.8 st) for “kuPAla“ type stress patterns. Duration also yielded relatively large differences (30 ms and 23 ms for first and second syllable, respectively). Intensity values differed less (about 1 and 4 dB in the first and second syllable, respectively) but nonetheless significantly.

Finally, spectral tilt also varied significantly as a function of prominence (see Table 2 for mean values). However, a 3-way ANOVA revealed, first, that this was only true when the second syllable of the pseudoword (i.e., the first syllable of the target word) was stressed (Position by Stress interaction, $F(1,58)=21.577$, $p < .001$), and the most notable changes in spectral tilt occurred in the highest frequency band (2-4 kHz) in the second syllable (Position by Stress by Frequency Band interaction,

$F(2,116)=15.564$, $p < .001$). In the first syllable, there was no significant difference in spectral tilt between stressed and unstressed vowels.

TABLE 1. Mean values of F0 (in Hz), intensity (in dB), and duration (in msec) of the first and second syllable of the test stimuli in Vroomen et al. (1998).

		Stress on the 1st syllable (e.g., KUpalo)	Stress on the 2nd syllable (e.g., kuPAlo)
1 st syllable	<i>F0</i>	112.7	104
	<i>Intensity</i>	79.7	78.1
	<i>Duration</i>	118	88
2 nd syllable	<i>F0</i>	90.2	122.9
	<i>Intensity</i>	78.0	82.3
	<i>Duration</i>	199	222

Table 2. Intensity (in dB) of the base (0-.5 kHz), low (.5-1 kHz), middle (1-2 kHz) and high (2-4 kHz) spectral frequency bands of the vowels on the first and second syllable of the test stimuli of Vroomen et al. (1998).

		Stress on the 1st syllable (e.g., KUpalo)	Stress on the 2nd syllable (e.g., kuPAlo)
1 st syllable	<i>0-.5 kHz</i>	79.8	78.7
	<i>.5-1 kHz</i>	76.5	75.3
	<i>1-2 kHz</i>	67.3	68.3
	<i>2-4 kHz</i>	69.7	69.5
2 nd syllable	<i>0-.5 kHz</i>	75.4	81.9
	<i>.5-1 kHz</i>	72.7	81.9
	<i>1-2 kHz</i>	66.1	76.0
	<i>2-4 kHz</i>	62.4	76.7

To analyze the contribution of the acoustic variables on reaction time speed, multiple linear regression analyses (Stepwise-method) were performed. Two separate models were created. The first model consisted of absolute values of the F0, intensity, duration, and the ratio of the spectral tilt (of the highest, and most significant, frequency band) of the first and the second syllable. The second model consisted of transformed difference scores of the F0 and intensity as a function of time (F01-F02 in semitones/sec, (dB1-dB2)/sec). In addition, the ratio variable of the duration ($\text{dur1}/(\text{dur1}+\text{dur2})$) and the overall spectral tilt measure of the vowel of

the second syllable was included in the model. In both models, the reaction time latency was the dependent variable.

In the first phase, we were interested in finding out what acoustic variables could explain the word spotting results when the reaction times were treated as a single set, that is, the model comprised all RTs from the two experiments as the dependent variable. These analyses showed that the only variable that predicted significantly the RTs was the F0. When absolute values were used as independent variables, the model explained about 66% of the total variance ($r^2=.664$, $F(2,117)=115.502$, $p < .001$). This model consisted of the F0s of the first and second syllable (F01: Beta=.230, $t=3.904$, $p < .001$; F02: Beta=-.692, $t=-11.736$, $p < .001$). When the F0 of the second syllable was entered alone in the model, power reduced only slightly, and the model explained 62% of the total variance ($r^2=.620$, $F(1,118)=192.520$, $p < .001$; Beta=-.787, $t=-13.875$, $p < .001$). When transformed variables were entered in the model, the F0 difference between the first and the second syllable (in semitones per second) turned out to be the only significant variable. The power of the model was about the same as with absolute values ($r^2=.671$, $F(1,118)=240.314$, $p < .001$; Beta=.819, $t=15.502$, $p < .001$).

In the second phase, regression analyses were performed separately for the two sets of RTs in order to find out if the same acoustic cues were employed when responses were fast or slow. The results revealed no significant correlation between any measured acoustic variables and RT within separate experiments. Correlation of the F0 measures with RT approached significance when reaction time latencies were fast but nonetheless were not included in the model.

DISCUSSION

The main finding of this study was that fundamental frequency was the only acoustic variable that significantly predicted reaction time speed in the two word spotting tasks in the study by Vroomen et al. (1998). This may be taken as indirect evidence that F0 is an important acoustic correlate of prominence in Finnish, and furthermore, that prominence is used as a cue to word boundaries.

It may be worthwhile to note that the excursion size of the F0 movement in our study was about 3 semitones. It is somewhat smaller than the ≥ 4 semitone difference required for effective pitch accent in Dutch ('t Hart et al., 1990). A second difference to Dutch pitch accent in the acoustic realization of the F0 movement was that spectral tilt seemed to be only a very weak correlate of prominence in Finnish. The most pronounced changes in spectral tilt were dependent on which syllable was

stressed, that is, spectral emphasis occurred only when the first syllable of the target word was stressed. A third difference in terms of how prominence was acoustically realized in pitch accented syllables in Dutch was that in our material the high frequency band (2-4 kHz) showed emphasis and not the middle band (Sluijter & van Heuven, 1996). Finally, duration plays a significant role in Dutch in the perception of prominence. In our study, durational differences were found between stressed and unstressed syllables, but they did not seem to contribute to the perception of prominence in any reliable way.

These differences may reflect language specificity in the acoustic realization of word level stress. In Dutch, the acoustic correlates of pitch accent are F0, duration and spectral tilt which all contribute to the perception of prominence. This prominence also signals linguistic distinctions. In Finnish, lexical stress may not have any linguistic significance, because it is highly predictable and in that sense redundant. However, lexically stressed syllables may still be more prominent for other than linguistic reasons. Prominence may be used as a cue to word boundaries, which in turn facilitates word recognition in continuous speech. Current results seem to suggest that acoustically prominence in our material was not realized as accent. In a separate study (see Chapter 4), we have also shown that sentence accent is not needed for fast reaction times in a word spotting task. Only a 2 semitone difference between the first and the second syllable seems to be sufficient to indicate prominence.

Finally, it should be noted the regression analyses yielded no significant results when the reaction times were analyzed within experiments (that is, fast and slow RTs separately). This might indicate that perceptually there are no intermediate prominence levels but listeners of Finnish make a binary distinction in terms of stressed vs. unstressed syllables.

To conclude, the results suggest that fundamental frequency is the most important acoustic correlate of prominence in this type of material. Other acoustic variables, such as duration or spectral tilt, which are reported to play a role in the perception of prominence in English and Dutch, did not predict reaction time speed. This finding may reflect language specific differences in the functional role of prominence. One possibility is that in Finnish prominence occurring on the lexically stressed syllables may not serve any linguistic function, but instead organizes the sensory input by creating a rhythm which in turn facilitates detection of word boundaries.

CHAPTER 4

WORD STRESS IN LEXICAL SEGMENTATION OF SPOKEN FINNISH¹

ABSTRACT

In Finnish, lexical stress (or primary word stress) is invariably located on the first syllable of a word. Recent evidence suggests that listeners are sensitive to word stress in detecting word boundaries in connected speech (Vroomen, Tuomainen, & de Gelder (1998) *Journal of Memory and Language*, 38, 133-149). However, there is dispute on how stress is realized acoustically in Finnish words. One possibility might be that only words with sentence accent have prominent initial syllables that help segmentation. If so, then prominence can be used only when the word receives sentence accent. With a word spotting task, we investigated whether listeners could easily detect words that were excised from a non-accented position in a sentence. The results were in close correspondence with previous findings suggesting that sentence accent is not needed. Thus, stress is an important cue to word boundaries in Finnish speech and is not dependent on sentence accent.

INTRODUCTION

Finnish is a language with fixed lexical stress on the first syllable of the word. Another typical phonological phenomenon of Finnish is (palatal) vowel harmony. In addition to being informative on the structure of word forms, both of these phenomena can provide cues to the boundaries of words in connected speech. This is particularly important because usually onsets and offsets of words in fluent speech are not reliably marked with cues such as pauses that clearly separate consecutive word forms. Instead, speech output consists of continuous flow of acoustic events that, in most instances, are not linearly organized with respect to the corresponding phonemes (Nygaard & Pisoni, 1995). Thus, listeners might look for any kind of systematic cues that would help segment the acoustic string into meaningful units.

¹ Chapter 4 is based on a manuscript by Tuomainen, J., Vroomen, J., & de Gelder, B. submitted for publication.

One way to conceive of the segmentation problem is that two different types of (interactive) processes are involved. First, current models of word recognition are based on the concept of lexical competition. Multiple lexical candidates activated by the same acoustic input compete (for example, through a process of lateral inhibition) until an optimal interpretation of the input is obtained, and the word is recognized. The boundaries of the word emerge as a by-product of recognition process. These types of models are SHORTLIST (Norris, 1990, 1994), and TRACE (McClelland & Elman, 1986). A second aspect of lexical segmentation concerns additional cues that help and speed up the segmentation process. Recent work by Cutler and her colleagues has revealed that often these cues are related to the phonological properties of specific languages, and thus are only useful for listeners who are familiar with those properties (e.g. Cutler & Norris, 1988; Cutler, Mehler, Norris, & Sequi, 1986; Cutler and Otake, 1994). A good example in English is the employment of the so-called Metrical Segmentation Strategy (MSS). The basic claim (which applies across different languages) is that word onsets usually align with the onsets of metrical units. In English, the metrical units are strong and weak syllables. Accordingly, every strong syllable (comprising a full vowel) is taken as the onset of the word (Cutler & Norris, 1988). This information can *bias* the activation process such that those lexical candidates that are aligned with the onset of the rhythmic unit (in this case, strong syllable) have a higher probability to survive in the competition (see Norris et al., 1995, for a computer simulation). Statistics of the structure of English support the plausibility of the strategy as about 90% of words start with a strong syllable (Cutler & Carter, 1987). Another structurally similar language to English is Dutch. In a similar vein, listeners of Dutch seem to employ the strong-weak metrical rhythm in lexical segmentation (Vroomen & de Gelder, 1995, Vroomen, de Zon, & de Gelder, 1996).

Potentially, lexical stress in Finnish is a very reliable cue to word boundaries because, in principle, all initial syllables are lexically stressed. As a consequence, initial syllables may stand out as more prominent than the neighboring syllables, and the most prominent syllable could be taken as the point for initiation of lexical access. Vowel harmony, on the other hand, refers to a phonotactic restriction on what types of vowels can occur in a noncompound word form. The Finnish vowel inventory comprises a total of eight vowels (/i y e ø æ u o ɑ/). Two classes of vowels, in terms of vowel harmony, can be distinguished, harmony vowels, and neutral vowels. Harmony vowels can further be classified as front (/y ø æ/) or back (/u o ɑ/) vowels. The neutral vowel set contains the vowels /i/ and /e/. The nature of the vowel in the first syllable of the word determines which vowels can follow in the non-initial syllable. In essence, if the first vowel belongs to one of the harmony sets, then only vowels from the same harmony set (or from the neutral set) can follow within a simple word form. For example, /juna/ ('train') and /hymy/ ('smile') are legal phoneme

strings, but */jyna/ or */junæ/ or */humy/ or */hymu/ are not (note: in Finnish orthography, /æ/ is written with <ä>, /ø/ with <ö>, and /ɑ/ with <a>).

Interestingly, listeners seem to be sensitive to these two types of phonological phenomena when they recognize words in connected speech. Suomi, McQueen, and Cutler (1997) showed that, in a word spotting task, participants were significantly faster in detecting target words when a vowel harmony mismatch indicated a word boundary as compared to a condition in which no mismatch was present (see Vroomen, Tuomainen, & de Gelder, (1998), Experiment 1 for replication). Thus, for example, /juna/ ('train') was detected faster in /PYjuna/ than in /PUjuna/ (the stressed syllables are written in upper case letters). However, in these experiments the acoustic realization of stress was not typical of Finnish in that the first syllable of the target was not the most prominent one. Instead, the prefix received the stress. To investigate the effect of stress as a cue to word boundaries, Vroomen et al. (1998) conducted another experiment in which the first syllable of the target word was stressed (for example, /puJUna/ or /pyJUna/). Two interesting observations emerged from this second experiment. First, response latencies were dramatically decreased by almost 500 ms as compared to Experiment 1 in which stress was mislocated. Second, no vowel harmony effect was found. Accordingly, Vroomen et al. (1998) proposed that stress was the primary cue to word boundaries and vowel harmony mismatch was only used as a secondary cue when stress provided incorrect information about the boundary.

The conclusion drawn by Vroomen et al. (1998) seems straightforward and intuitive. Fixed word stress on the first syllable marks, in principle, every word boundary. Vowel harmony is less effective because only a harmony mismatch indicates word boundary. If no harmony mismatch occurs there is no way of telling (on the basis of vowel harmony) whether a boundary has occurred or not. However, one important concern regarding the generalizability of the results of Vroomen et al. is that they did not explicitly state how stress (or prominence) was acoustically realized in their stimuli. It is possible that stress was realized as sentence accent so that prominent syllables stood out more clearly than they would if prominence were realized as is typical in other lexically stressed syllables. If sentence accent was used to indicate syllable prominence in Vroomen et al., it could weaken the claim that, in Finnish, word stress is used as an online cue to word boundaries. Location of sentence accent is determined on the basis of sentence structure, and only a small number of words receive accent. Thus, stress could then not be a commonly available cue to word boundaries. If indeed sentence accent was used in their experiment, that fact could also in part explain the extremely fast response speed in the second experiment of Vroomen et al. (1998). The response latencies were 500 ms faster than those in the first experiment where stress provided incorrect information about the word boundary. A well-documented finding in the psycholinguistic literature is that in a variety of tasks RTs are faster to stimuli with sentence accent (for a recent review, see

Cutler, Dahan, & Donselaar, 1997). One recurring explanation for faster RTs is that accented syllables are perceptually more salient and easier to process. Among other things, facilitation is related to more pronounced differences in the acoustic parameters between accented and non-accented syllables although acoustic clarity is not the whole story (Cutler, 1976).

Thus, the major motivation for the present study was to investigate whether sentence accent is needed for fast reaction times when listeners need to segment continuous speech input. The other alternative is that less prominent acoustic changes (more typical of lexically stressed unaccented syllables) are sufficient. It should be noted that the relationship between lexical stress and its acoustic realization in Finnish might be problematic. The main reason is that since lexical stress is fixed, it is completely predictable and there may be no need for overt marking, at least for linguistic purposes. Unfortunately, to date, relevant research on the acoustic correlates of word stress in Finnish is lacking (see Iivonen, (1998) for a recent account of Finnish prosody including some details of word stress and sentence accent). Acoustic correlates of stress and accent are fundamental frequency (F_0), intensity, duration, and vowel quality (Fry, 1958). There is some, although indirect evidence that, in Finnish, F_0 is one of the important acoustic correlates of stress/accent, at least in this type of stimuli. Tuomainen et al. (1999) performed acoustic analyses of the stimuli used by Vroomen et al. (1998). The reaction times obtained by Vroomen et al. were then regressed to the acoustic measures. A model consisting of F_0 , duration, intensity, and spectral tilt values was constructed. The results showed that the F_0 was the only acoustic parameter that significantly predicted reaction times in the two experiments by Vroomen et al. (1998). The F_0 difference between the first and the second syllable in their second experiment, in which the stress was on the second syllable of the nonsense string (i.e., /puJUna/), was 2.8 semitones (18 Hz). Whether this difference is typical of word level stress or sentence accent in Finnish is at present unknown. If a comparison is made to other languages, the difference is smaller than a corresponding F_0 difference, for example, in Dutch, where at least a 4 semitone difference is typically required for sentence accent ('t Hart et al., 1990). Of particular importance regarding other acoustic correlates of stress is the role of duration. In English and in Dutch, both so-called stress-timed languages, duration co-varies with stress, so that stressed syllables are longer than unstressed syllables. Even though durational differences were found by Vroomen et al, and duration co-varied with stress, duration did not significantly predict reaction time speed. One possibility for this may be that length (or phoneme quantity) is a distinctive feature of the Finnish phonological system. Vowels in all syllable position can be phonologically short or long. As a consequence, the vowel of a stressed syllable can be (both acoustically and phonemically) the shortest of the word, as is exemplified in a word such as *kanuuna* ('*canon+part*'). However, to be sure, the role of duration as an acoustic cue to stress in Finnish needs to be tested experimentally.

Finally, an important detail concerning the acoustic realization of lexical stress in Finnish is that no (significant) vowel reduction takes place in the unstressed syllable, a typical feature, for example, of English. Thus, in Finnish, vowel quality (or 'coloring') is not informative of the syllable's stress level.

Because of the scarcity of the acoustic details on how sentence accent and lexical stress are realized in Finnish, we set up an experiment in which participants responded to target words that were excised from a *non-accented* position in a sentence. This approach allows us to investigate the segmentation process without knowing the exact details of the acoustic correlates of sentence accent or lexical stress. The predictions of the current experiment were straightforward. If RTs were comparable to the second experiment of Vroomen et al. (1998) (that is, fast RTs when stress was located on the first syllable of the target, and the absence of a vowel harmony effect), then this suggests that sentence accent is not required to speed up segmentation. However, if the RTs were slow (and comparable to the first experiment by Vroomen et al.) then this would suggest that the results of Vroomen et al. were a special case limited to words with sentence accent. In the latter case, similar to the findings by Suomi et al. (1997) and Vroomen et al. (1998), one would also expect to find a vowel harmony effect (i.e., faster RTs to targets preceded by a disharmonious prefix compared to targets preceded by a harmonious prefix).

EXPERIMENT

METHODS

Participants. Twenty-two native speakers of Finnish with no known auditory deficiencies received a small fee for participation. They were students from the University of Turku or Åbo Akademi University.

Materials. A total of eighty-eight target words were selected from a massive unpublished computerized corpus of contemporary written Finnish (22.7. million word tokens; Laine & Virtanen, 1999). All words were monomorphemic CVCV nouns or adjectives. Forty-four were high-frequency words ($> 36.1/\text{mil}$, mean 275.7 s.d. 347.9) and forty-four were low-frequency words ($< 5.3/\text{million}$, mean 1.6, s.d. 1.21). Eighty-eight CVCV nonsense filler items were also constructed by changing one or two phonemes of the experimental real word targets. All targets and fillers were embedded in a sentence context in a non-accented position. They were spoken by one of the authors (JT; note: he also produced the stimuli in Vroomen et al., 1998), and recorded on a DAT tape. In all cases, sentence accent was located on the target preceding word. As an example, the target word *salo* ('backwoods') was pronounced in a sentence frame /sano

h*alpa salo kaksi kertaa/ ('say ch*eap backwoods two times [and not expensive backwoods]). In this example, the contrastive sentence accent was on the first syllable of the word /halpa/ ("cheap") (see Figure 1 for an example of the acoustic realization of *F0* and intensity). *F0* values were measured at four positions: The syllable preceding the one that received sentence accent; the accent receiving syllable; the syllable before the target initial word; and the initial syllable of the target word itself (the arrows in Figure 1 indicate the critical syllables). Results showed that, on the average, the *F0* of the accented syllable differed from the *F0* of the preceding unstressed syllable by 3.1 semitones (19 Hz) (e.g., /sa-no h*al-pa/; the critical syllables are underlined). In the non-accented position, the *F0* of the stressed syllable was on the average 1.9 semitones (10 Hz) higher than the *F0* of the preceding unstressed syllable. Note that the *F0* difference in the stimulus material of Vroomen et al. (1998) was 2.8 semitones (18 Hz), which is almost the same as in the accented syllable in the current material. This suggests that in Vroomen et al. an *F0* contour typical of sentence accent might have been employed, although as already mentioned earlier, this difference is smaller than the one required for sentence accent for example in Dutch ('t Hart et al., 1990). The sentences were digitized at 22050 Hz (16 bit resolution) and saved on the hard disk of a PC. The target words were excised from the sentence, and later a CV syllable (excised from an unstressed position preceding the target word) from the same set of sentences was spliced into the target word creating a CVCVCV nonsense string. The CV syllable could either be harmonious (e.g., /pasalo/) or disharmonious (e.g., /pæsalo/) with the target word. A similar procedure was also applied to the pseudoword fillers. Vowel height was kept constant, that is, /u/ was always paired with /y/, /o/ with /ø/, and /a/ with /æ/. The splicing technique was used so that the identical token of the target could be used in both conditions, and possible acoustic differences between the same target word in different conditions would not be a confound. Special effort was taken to ensure that none of the CVCVCV nonsense strings contained an embedded real word. In addition, the uniqueness points and cohort sizes for real words and deviation points for pseudoword fillers were computed. These features have an effect on recognition latencies via lexical competition. More specifically, differences in the uniqueness point might affect the frequency effect because different words were used. Furthermore, if the deviation point of the nonword fillers was always located early in the nonword, then this could lead to a response strategy which could vanish the vowel harmony effect. Finally, the possible cohort size differences relate to the vowel harmony effect. Only in the harmonious targets the first two syllables of the nonsense string may form a real word. The larger the cohort size the slower the recognition latency will be, and as a consequence, the vowel harmony effect may reflect lexical competition, and not the utilization of vowel harmony mismatch as a cue to word boundary (see Suomi et al., 1997, for a similar concern).

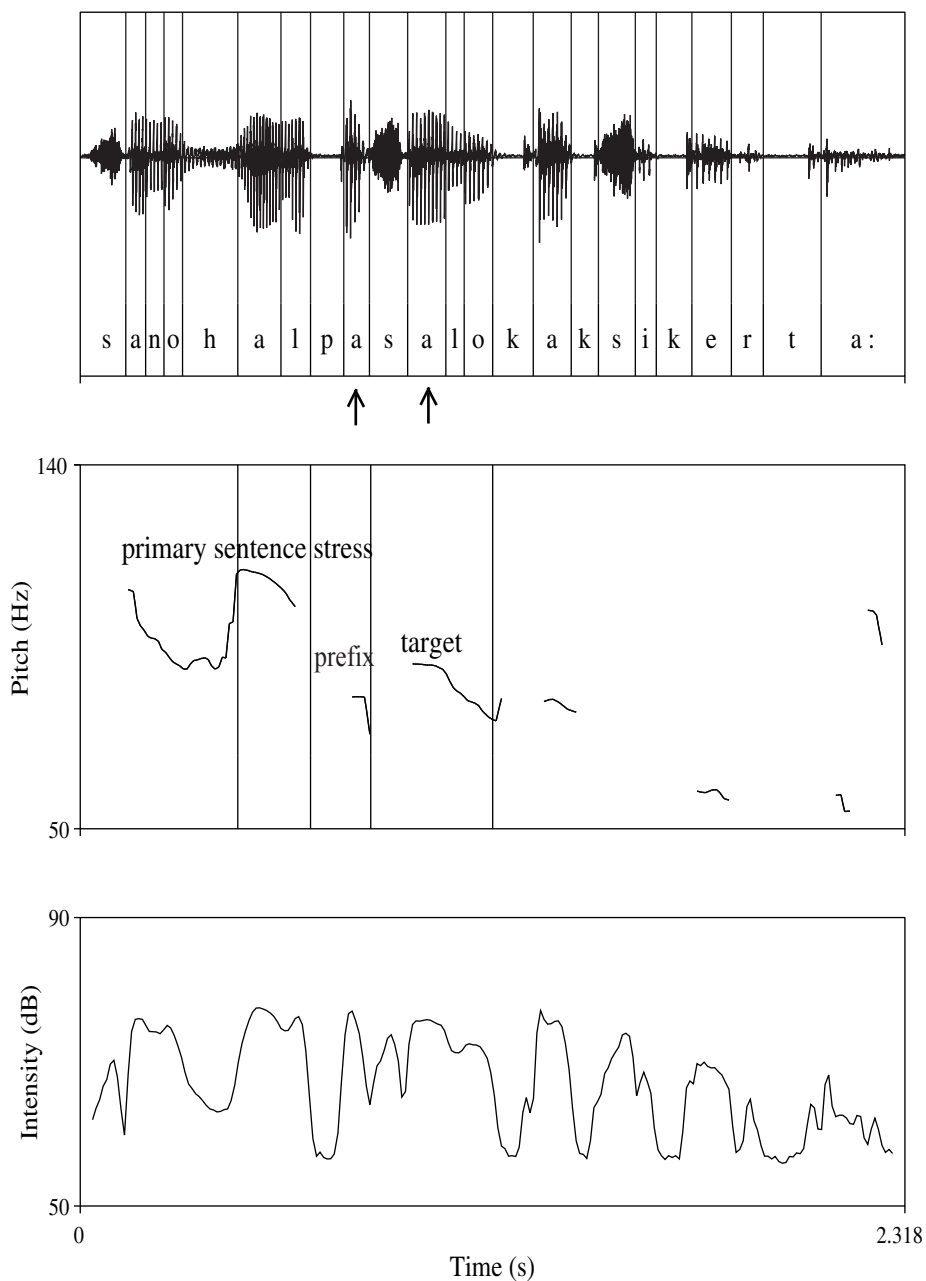


FIGURE 1. Examples of some of the acoustic parameters (waveform, intensity curve (in dB), and F0 curve (in Hz)) of a sentence /sano h*alpa salo kaksi kertaa/ ('say cheap backwoods two times'). The target word (*salo*) and the prefix (in this case *pa*) used in the experiment were excised from a non-accented position (indicated by an arrow).

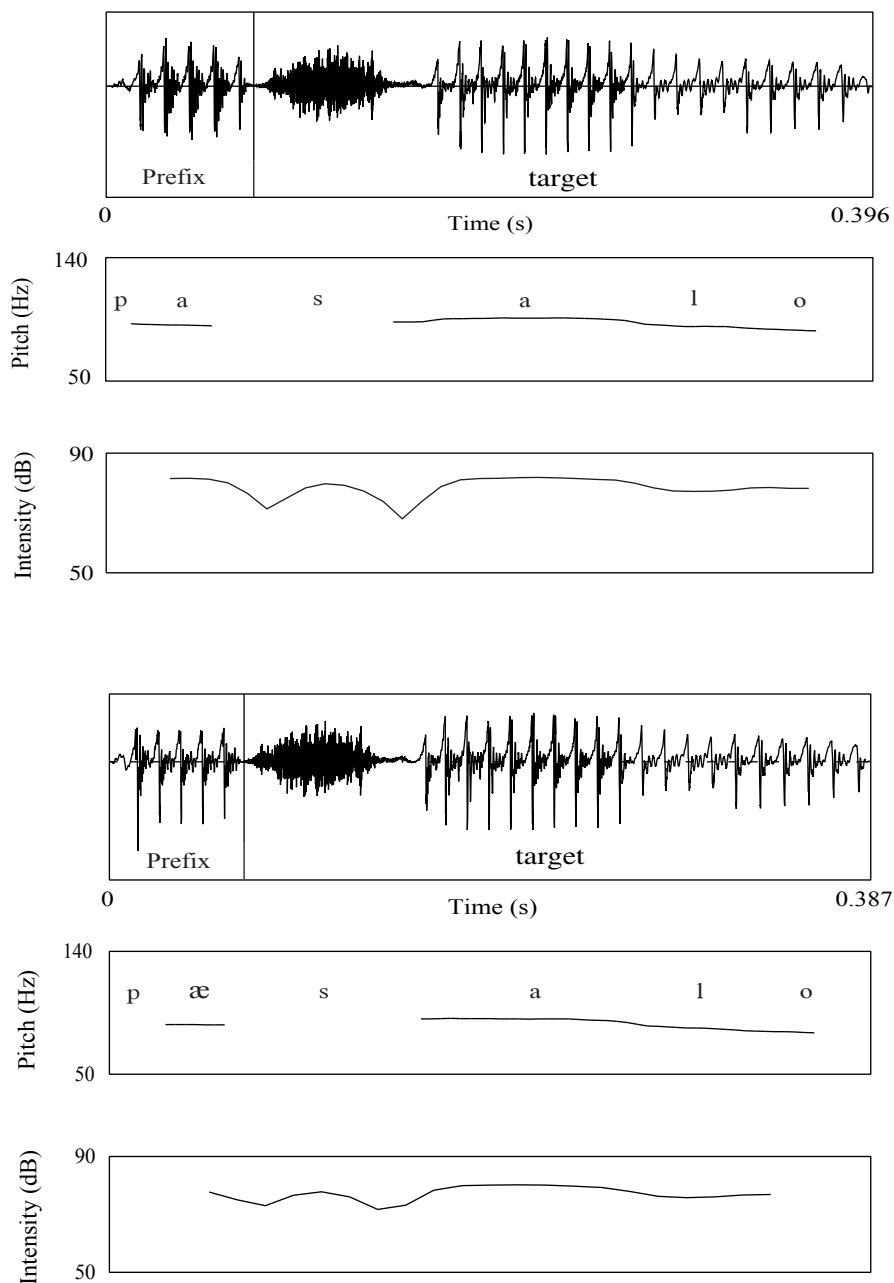


FIGURE 2. Example of the target word that was spliced into a harmonious (upper panel) and disharmonious (lower panel) CV prefix. To control for any possible acoustic flaws introduced by splicing, the stimuli were re-synthesized and the pitch difference between the prefix and the first syllable of the target was adjusted to two semitones (according to the difference in the sentence materials) by the PIOLA technique (a modified PSOLA method).

In the current material, all real word targets were unique either on the final phoneme (37% of targets) or after the offset of the word. Two per cent of the pseudoword fillers deviated from real words on the 2nd, 12% on the 3rd, 55% on the final phoneme, and 31% after the offset. These figures indicate that most of the stimuli were unique on the final phoneme or after the offset which suggests that participants needed to listen the whole nonsense string and could not rely on a specific strategy in order to response accurately. 25 % of the high frequency targets, and 50% of the low frequency targets were unique on the final phoneme (the rest were unique after offset). This difference is significant ($\chi^2_{(1)} = 13.33$, $p < .001$). However, it suggests that low frequency targets could be recognized earlier, and thus decrease the possible frequency effect.

55 % of the high frequency targets and 41% of the low frequency targets had one or more competitors (that is, longer words that were consistent with the first two syllables of the nonsense string). This difference in the distribution is not significant ($\chi^2_{(1)} = 1.639$, n.s.). More importantly, 47% of the harmonious targets were consistent with one or more longer words. Thus, if a vowel harmony effect was obtained with the current material it might be due to lexical competition and not to vowel harmony mismatch. Because of the fairly stringent stimulus selection criteria we could not replace harmonious items without affecting, for example, the frequency counts. As will be noted later in the Results section, lexical competition is not an issue regarding present results.

To control that the splicing did not introduce acoustic artifacts, the experimental stimuli and fillers were resynthesized using the PIOLA technique (Pitch Inflected Overlap Add), which is a modification of PSOLA (Moulines & Charpentier, 1990) implemented in the GIPOS speech analysis software at the Center for Research on User-System Interaction (IPO, Eindhoven). At the same time, the F_0 contour between the prefix and the target initial syllable was stylized and the F_0 difference was controlled so that in all stimuli the F_0 of the prefix was two semitones lower than the F_0 of the first syllable of the target (see Figure 2 for an example of the resulting waveform, F_0 and intensity contours of the harmonious and disharmonious items). The duration and amplitude of the prefix (and the targets) were kept constant across conditions. The resulting stimuli sounded natural and no acoustic artifacts could be noted².

² To test that splicing did not introduce any audible artifacts, we ran a listening task in which five participants evaluated whether a stimulus was a natural (unmodified) CVCVCV string or a synthesized stimulus. Participants were staff members or graduate students of the Centre for Cognitive Neuroscience (of which one had phonetic training and experience with synthetic speech). The stimuli were 22 natural CVCVCV tokens excised from the original sentences,

Design and procedure. As in Vroomen et al. (1998), two lists of stimuli were constructed so that a participant heard each embedded target word (either with a harmonious or a disharmonious prefix), and only once. The type of prefix was counterbalanced across lists. The position of the fillers and each member of an experimental item pair was the same in the two lists. The interstimulus interval was 4.5 seconds. A practice list of 24 trials preceded the experiment. Participants were tested individually in a quiet room. Stimuli were presented through a loudspeaker, and a PC controlled the stimulus presentation and data collection. The task of the participant was to listen to nonsense items, which sometimes contained a finally embedded real word (word spotting, McQueen, 1996). They were instructed to press a response button with their preferred index finger as soon as they heard a real word, and after that say the word aloud. The vocal responses were checked by the experimenter in order to determine whether the intended word was detected.

RESULTS

Mean response times (RTs) and miss rates (i.e., no response on a target) were computed. Response times were measured from the onset of the embedded target, and vocal responses that did not correspond to the intended word were treated as errors (0.2%). Outlying responses that were slower or faster than 2.5 standard deviations from the individual subject or item means (3% of all responses) were also regarded as errors. Inspection of individual items indicated that three (low frequency disharmonious) items (*rysä* 'fyke', *jyly* 'rumble', *kumu* 'din') yielded error rates higher than 60% and were discarded from further analysis together with the related member of the item pair. None of the participants had an error rate higher than 40%, and none were therefore excluded. The false alarm rate was 2.4%. Analyses of variance (ANOVA) were performed with subjects (*F1*) and items (*F2*) as random

and 22 were experimental (spliced and synthesized) stimuli. These specific stimuli were chosen so that the first consonant of the target word was not a plosive. In this way, a more stringent test of the stimulus quality could be performed as splicing a CV prefix into a target word beginning with a voiced consonant is the most likely combination to create audible artifacts, for example, due to coarticulation mismatch or unsuccessful matching of periodic cycles in the two acoustic samples. The stimuli were presented one at a time with an ISI of 3 seconds. The participants responded by pressing one of two buttons marked with a label "natural" or "synthetic". The results showed that, in general, participants identified 53% of the items correctly. According to a binomial test (with $p < .05$) performance at or above 66% in a 44-item test is significantly better than chance. None of the participants scored better than 61% correct. Error analysis indicated that participants correctly identified 80% of the natural stimuli as natural, and only 20% of the synthetic as synthetic. Thus, they were biased towards perceiving the synthetic stimuli as natural stimuli (the difference in the distribution is significant according to a χ^2 -test ($\chi^2_{(1)} = 60.65$, $p < .001$).

variables. In the subject analysis, frequency of occurrence of the target word (high vs. low) and prefix type (harmonious vs. disharmonious) were within-subject factors. In the item analysis, prefix type was a within-items factor and frequency was a between-items factor.

TABLE 1. Reaction time (RT) latencies (in ms) and error rates (in parentheses) to the high and low frequency target words (upper panel). Results from Vroomen et al. (1998; Experiment 1 on the left, and Experiment 2 on the right) are presented in the lower panel. RTs are measured from the onset of the target word.

EXPERIMENT (F0 PEAK ON THE 2 ND SYLLABLE)			
PREFIX TYPE	TARGET TYPE		(mean)
	<i>High frequency</i>	<i>Low frequency</i>	
harmonious	687 (4%)	755 (11%)	720 (7%)
disharmonious	686 (5%)	736 (11%)	711 (8%)

VROOMEN ET AL. (1998), EXPERIMENT 1 (F0 PEAK ON THE 1 ST SYLLABLE) AND 2 (F0 PEAK ON THE 2 ND SYLLABLE)		
PREFIX TYPE	EXPERIMENT 1	EXPERIMENT 2
harmonious	1217 (19%)	704 (7%)
disharmonious	1082 (11%)	690 (7%)

The results are presented in Table 1. For convenience of comparison, RTs from the two experiments of Vroomen et al. (1998) are included in the Table. The RTs from Vroomen et al. are collapsed over front and back vowel items. As can be seen, results were comparable to the ones obtained by Vroomen et al. (1998) in their second experiment in two respects. First, overall latencies were almost identical, second, there was no harmony effect.

Further inspection of the Table shows that high frequency targets were responded to 68 ms faster than low frequency targets ($F(1,21) = 58.42$, $p < .001$, $F(1,83) = 9.44$, $p = .003$).

The other important aspect of the data is the absence of a vowel harmony effect, a similar finding to Vroomen et al. (1998) in Experiment 2. In the current experiment the targets preceded by a disharmonious prefix were detected 10 ms earlier than those with a harmonious prefix, while in Vroomen et al. the difference was 14 ms. The 10 ms difference turned out to be significant in the participant but not in the item analysis ($F1(1,21) = 5.01$, $p = .036$, $F2(1,83) = 1.02$, n.s.) (as was also the case in Vroomen et al.). In Vroomen et al. (Experiment 1) the size of the vowel harmony effect was 135 ms, which is over ten times larger than the effect in the present study. Furthermore, no interaction was found.

To ensure that there was no harmony effect as a function of frequency, we further analyzed the RTs separately for the high and low frequency stimuli. Comparison of mean RTs indicated that the difference between harmonious and disharmonious stimuli was not significant either for the high frequency items ($t1(21) = 0.105$, n.s., $t2(43) = -0.011$, ns) or for the low frequency items ($t1(21) = 1.648$, n.s., $t2(40) = 1.062$, n.s.).

In order to explore the possibility that stimulus duration might have played a role in the vowel harmony effect, we measured the reaction times also from the offset of the nonsense string. This measure corresponds closely to the uniqueness point of the target words of the current stimulus material as 63% of the targets were unique after the offset, and the rest on the final phoneme. Mean RTs were 384 ms for the high frequency harmonious, and 381 ms for the disharmonious targets. The corresponding values for the low frequency targets were 465 ms and 435 ms. By participant and by item analyses (repeated measures ANOVA) indicated a main effect of frequency ($F1(1,21) = 77.52$, $p < .001$, $F2(1,82) = 13.08$, $p = .001$). Furthermore, a main effect of vowel harmony was also present, but it was significant only in the participant analysis ($F1(1,21) = 14.33$, $p = .001$; $F2(1,82) = 1.73$, n.s.). The interaction between frequency and vowel harmony was not significant ($F1(1,21) = 2.64$, $p = .119$; $F2(1,82) = 1.32$, n.s.), but we still performed planned comparisons separately for the high and low frequency harmonious and disharmonious targets. These analyses revealed that the 3 ms difference in high frequency items was not statistically significant ($t1(21) = .346$, n.s., $t2(42) = .232$, n.s.). The low frequency disharmonious targets were detected 30 ms faster than the harmonious targets, but this effect was only significant in the participant analysis ($t1(21) = 3.255$, $p = .004$; $t2(40) = 1.294$, n.s.). Overall these results are in close correspondence with the ones performed on RTs measured from the onset of the target word.

Analysis of the error rates indicated only a significant frequency effect ($F(1,21) = 20.68$, $p < .001$, $F(1,82) = 9.24$, $p < .003$). Harmony class of the prefix neither showed a significant main effect nor interacted with frequency.

Next we compared the overall RTs of the present Experiment to the RTs of the two experiments by Vroomen et al. (1998). To ensure that items were balanced by frequency of occurrence, we first computed a frequency count of the items employed by Vroomen et al. using the Turun Sanomat corpus (Laine & Virtanen, 1999). The items of Vroomen et al. had a mean frequency of occurrence of 57.1/million, which falls between the frequencies of our high (mean 275.7) and low (mean 1.2) frequency items. A (non-parametric) t-test showed that the difference was significant both when compared to the high frequency items ($U = 140$, $p < .001$) or the low frequency items ($U = 40$, $p < .001$). As a consequence, we decided to compare the RTs obtained by Vroomen et al. to the overall RT collapsed over the high and low frequency targets of the current experiment (mean frequency 138.7/million, which is not significantly different from Vroomen et al. [$U = 1088$, $p = .509$]). It can be seen from the Table that the mean RTs in the current experiment are in close correspondence with the RTs in the second experiment of Vroomen et al. (1998). The mean value in our experiment was 716 ms as compared to 697 ms in Vroomen et al., resulting in a 19 ms difference. Also, the error rate in the current experiment (5%) is comparable to the second experiment of Vroomen et al. (7%).

Taken together, our results are in close correspondence with the ones obtained by Vroomen et al. (1998) in their second experiment in two respects. First, reaction time latencies are practically identical (in the range of 20 ms). Second, no reliable harmony effect was noted.

DISCUSSION

We investigated the role of word stress and vowel harmony mismatch as cues to word boundaries in Finnish. The main motivation for the study was the uncertainty about the generalizability of the results by Vroomen et al. (1998) who claimed that word stress is more important in lexical segmentation than vowel harmony mismatch. The major concern was the acoustic realization of stress in their stimuli: were the word-initial syllables perceived as if they carried sentence accent or not?

The present results showed that listeners were able to spot target words excised from non-accented position with speed and accuracy that was comparable to the results of Vroomen et al. (1998) in their second experiment. A second

commonality is that we failed to obtain a reliable vowel harmony effect. Disharmonious targets were detected some 10 ms faster than harmonious targets but this result was only significant in the analysis when participants were used as a random variable. One may also note that there is some indication that the difference in the response speed between harmonious and disharmonious targets was larger for the low frequency items, which were responded to slower than the high frequency targets. This may suggest that detection of vowel harmony mismatch takes considerable amount of time and the harmony effect will be more pronounced the slower the response speed.

Our conclusion here is that sentence accent is not needed for fast reaction times. We suggest that word level stress (or prominence) may be signaled by small acoustic differences (in the current case by a two semitone difference in the F_0). These results support the idea that word stress is an important cue to word boundary in Finnish independent of sentence accent.

Our results also suggest that word stress can be used as an online cue in word recognition. This implies that the prominence level of a syllable can be computed fast, and is not a time consuming operation as suggested by, for example, Suomi et al. (1997). A traditional view is that stress (and other suprasegmental features) are relative properties of syllables (Fry, 1958). Accordingly, before knowing whether a particular syllable is stressed, one needs to compare it to adjacent syllables, which in turn implies an analysis window of several hundreds of milliseconds. However, an alternative view to this could be that stress might be indicated by more paradigmatic properties of a syllable. These types of parameters are, for example, typical F_0 contour on a syllable, intensity or more accurately the spectral properties (that is, increased amplitude of higher harmonics; Sluijter & van Heuven, 1996), which may be computed without reference to surrounding syllables. To be sure, at least for F_0 , an integration time is needed (Hermes, 1997) which is in part reflected by the fact that recognition of a typical pitch contour on an accented syllable is slower than recognition of vowel quality of the same syllable (e.g., Cutler & Chen, 1997). However, the integration time probably spans only one syllable, which is short enough for this type of cue to qualify as an online cue in word recognition.

As already mentioned in the introduction, the acoustic correlates of lexical stress in Finnish are currently not known. However, there is indirect evidence at least from two sources that suggest that a rise in F_0 is one way of implementing word stress and sentence accent in spoken Finnish. First, Iivonen, Niemi, and Paananen (1998) investigated the occurrence of F_0 peaks on lexically stressed syllables in Finnish, German and English radio and TV news broadcasts. The results showed

that there were fewer *F0* peaks (above the one semitone criterion set by Iivonen et al.) in Finnish than in German or English, probably corroborating the impression of speakers of Indo-European languages that the "melody" of Finnish "sounds flat". However, 73% of all *F0* peaks in Finnish occurred on the lexically stressed syllables (that is, on the first syllable of the word). The corresponding figures in German and English were 42% and 59%, respectively. Finally, 52% of all lexically stressed syllables as opposed to 27% in German, and 40% in English contained an *F0* peak.

Second, as already mentioned in the Introduction, Tuomainen, Vroomen and de Gelder (1999) ran acoustic analyses on the stimuli employed by Vroomen et al. (1998). They measured the *F0*, duration, intensity and spectral tilt of (the vowels) of the first and second syllable of the nonsense carriers (i.e. the prefix and the first syllable of the target). Subsequent regression analyses indicated that only *F0* (and not duration, overall intensity or spectral tilt) significantly predicted reaction time speed in the two word spotting experiments by Vroomen et al.

When one considers the phonological structure of Finnish, the assumption that *F0* is important in signaling prominence seems reasonable. One typical feature of Finnish is the use of length as a distinctive feature. Thus, increasing (i.e., doubling) the duration of the first vowel [a] in /hara/ ('harrow') changes the word to /haara/ ('branch'). In the former, the duration of the vowel of the first syllable is almost 1.5 time shorter than the vowel in the second syllable (Lehtonen, 1970). As a consequence, durational differences need to be carefully controlled so that the meaning of the word does not change (see also Suomi et al., 1997 for a similar view). Another acoustic parameter, intensity, refers to the increased pulmonary effort during the production of stressed syllables. In general, overall intensity correlates poorly with perceived stress. One of the reasons is that intrinsic variation as a function of vowel identity reduces the value of intensity as a reliable cue to prominence. Low vowels are more intense than high vowels. However, in Finnish, both can occur without reduction in stressed and unstressed syllables. Another reason, as Sluijter and van Heuven (1996) suggest, is that intensity is too susceptible to noise. Accidental events, such as the speaker turning his head, will cause significant intensity drops, which are much larger than those caused by the difference between stressed and unstressed syllables.

Thus, although current research on this issue provides only indirect evidence as to the acoustic correlates of word stress in Finnish, it points to the direction that *F0* is at least one important acoustic correlate of prominence in Finnish. However, in this respect, the ubiquitous, though irksome statement – further studies with a larger set of speakers and stimulus material are necessary – appears appropriate.

To summarize, our results indicate that sentence accent is not needed for fast responses. As already suggested by Vroomen et al. (1998), word stress seems to be an important cue to word boundaries in Finnish. Finally, since we found no reliable effects of vowel harmony either on the reaction time speed or on the accuracy of performance, our results support the view, offered by Vroomen et al., that word stress is the primary cue that listeners employ in detecting word boundaries in spoken Finnish.

CHAPTER 5

DIFFERENT TEMPORAL DYNAMICS IN THE UTILIZATION OF WORD STRESS AND VOWEL HARMONY IN LEXICAL SEGMENTATION OF FINNISH: AN EVENT-RELATED BRAIN POTENTIAL STUDY¹

ABSTRACT

In Finnish, lexical stress is located invariably on the first syllable of the word. Fixed stress provides a cue to word boundary, and facilitates recognition of words in continuous speech. However, if stress is mislocated, listeners attend to other cues, such as vowel harmony mismatch, which also speeds up (but to a lesser degree) lexical segmentation. In four experiments, we investigated the temporal dynamics of multiple cues to word boundaries in spoken Finnish by measuring both reaction times and event-related potentials (ERPs). The behavioral results showed that word stress greatly facilitated recognition speed of both real word and pseudoword targets, and that a vowel harmony effect was obtained (for both real and pseudoword targets) only when the stress cue provided incorrect information about the boundary. Correct stress position speeded up pseudoword detection more than detection of real words. ERPs revealed that the facilitatory effects of stress relates mostly to early, possibly pre-lexical detection of stressed syllable as indicated by an early negativity (between 400 and 600 ms from the onset of the CVCVCV carrier item) both in the left and right hemispheres. The detection of vowel harmony mismatch began simultaneously with the word recognition but extended notably past the acoustic offset of the target. Early involvement was most pronounced in the left-sided electrodes, and the ERP response indicating late processing was more wide spread in the parietal areas. In addition, only weak responses were obtained for pseudoword targets in the early time window. This suggests that for real words the computation of the harmony mismatch may involve the phonological code related to the word forms. Current results corroborate the findings by Vroomen, Tuomainen, and de Gelder (1998) in that stress seems to be the primary and vowel harmony mismatch the secondary cue in lexical segmentation of Finnish. The temporal characteristics of the ERP waveforms suggest that this is due to the earlier availability of the stress cue.

¹ This chapter is based on a manuscript with the same title by Jyrki Tuomainen, Koen Böcker, Jean Vroomen, and Beatrice de Gelder.

INTRODUCTION

Word boundaries in spoken language are not indicated by clear and consistent cues. Typically, in this respect, spoken language is contrasted with written language in which white spaces between words reliably show where one word begins and the preceding word ends. Thus, if spoken language were like written language, spoken words would be separated by pauses or other invariant acoustic events. However, despite the lack of clear cues to word boundaries, human listeners show no difficulty in perceiving and experiencing speech as consisting of a flow of consecutive words.

How is this seemingly effortless goal accomplished? Several alternative models have been suggested on how word boundaries are detected. A major distinction can be made between models that suggest that word boundaries emerge as a by-product of word recognition (e.g., Marslen-Wilson, 1984, McClelland & Elman, 1986, Norris, 1994), and those that assume explicit marking of word boundaries (e.g., Cutler & Norris, 1988). In lexical models, the acoustic signal is converted into a string of phonological units that are continuously mapped onto lexical representations of words. Models differ regarding the exact nature of lexical processing. In TRACE (McClelland & Elman, 1986) and Shortlist (Norris, 1994), but not in the Cohort model (e.g., Marslen-Wilson, 1984), lexical candidates (i.e., the word form representations that match the acoustic input) compete with each other for selection at the lexical level. As more phonemes are revealed in the input those candidates that are incompatible with the input will drop out of the competition. Finally, when one of the candidates has received sufficient amount of activation, the word is recognized, and word boundaries are automatically revealed (e.g., McQueen, Norris, Cutler, 1994).

Other cues to word boundaries are also employed by listeners. One source of information in the speech signal is provided by the prosodic features of a language. The cues that listeners are sensitive to differ as a function of the metrical structure of the specific language. For example, in English, the alternation of strong and weak syllables creates a rhythm that is exploited by listeners in detecting word boundaries. A strong syllable contains a full vowel as opposed to a reduced vowel (usually 'schwa') in a weak syllable. When a strong syllable is encountered, a word boundary is postulated (e.g. Cutler and Norris, 1988). This so-called Metrical Stress Strategy (MSS) is reasonable since most of the words in English begin with a strong syllable (Cutler & Carter, 1987). Another language with similar metrical structure is Dutch in which MSS has also been shown to be effective (Vroomen & de Gelder, 1995). However, the MSS is not useful for French listeners, because there is no (significant) vowel reduction as a function of stress in French. However, French

listeners seem to pick up the rhythm created by syllables (Cutler, Mehler, Norris, & Seguí, 1986). As it happens, for a native speaker of Japanese, the syllable structure per se does not facilitate recognition of words, but instead, a rhythmic unit corresponding to a subsyllabic unit, 'mora', seems to help segmentation (Otake, Hatano, Cutler, & Mehler, 1993). All these results provide evidence that the metrical structure of a language is one important source of information that listeners can use in lexical segmentation.

Another example of additional cues to word boundaries comes from Finnish, a Finno-Ugric language spoken by some 5 million people mostly in Finland. Finnish is a language in which lexical stress is invariably located on the first syllable of the word. Languages with fixed stress are fairly common. Hyman (1977) surveyed 444 languages of which, according to his analysis, about two thirds have fixed primary word stress. Furthermore, the analysis suggested that in about 70% of languages with fixed stress position, the primary word stress was dominantly located either on the first or the last syllable. As such, fixed stress is a potentially reliable cue to a word boundary. Thus, listeners of Finnish can assume with great certainty that every time a stressed syllable is encountered, it is the first syllable of a word. Another potential cue to word boundary is vowel disharmony. Vowel harmony refers to a phonotactic rule, which restricts the occurrence of vowels within a (native) word form. Accordingly, the quality of the leftmost vowel in the word stem defines what other vowels can occur later in the same word. The palatal vowel harmony of Finnish restricts the occurrence of vowels belonging to front or back harmony set in that only vowels from one set (and so-called neutral vowels, /i/ and /e/) can occur within a word. Thus, *juna* ('train') or *hämý* ('twilight') are legal words, but *jyna* or *junä*, or *hamy* or *hämu* are not. When vowels in the adjacent syllables belong to the other harmony set, a word boundary can be assumed. It should be noted that only a harmony mismatch informs about a boundary. If consecutive words contain only vowels belonging to same harmony set, it is not possible to determine on the basis of vowel harmony whether or not a boundary occurred (as in *Jukka juoksee*, 'Jukka runs').

Both word stress and vowel harmony have been shown to have functional consequences in lexical segmentation of spoken Finnish. In a word spotting task, Suomi, McQueen, & Cutler (1997) demonstrated that a target word *katu* ('street') was detected faster in /py.katu/ than in /pu.katu/. Presumably the disharmonious prefix indicated a word boundary between /py/ and /ka/, which facilitated word recognition. A similar finding was also obtained by Vroomen, Tuomainen, & de Gelder (1998). Furthermore, in two separate experiments, using different types of behavioral tasks, Vroomen et al. also demonstrated that word stress was an effective cue to word boundary and also specific to the Finnish language. Thus, when the first

syllable of the target word was stressed (or more prominent than the surrounding syllables), the Finnish participants benefited from word stress more than, for example, the Dutch or French listeners.

What happens if listeners have multiple and, in some instances, conflicting cues to word boundaries available at the same time? For example, in Finnish, not infrequently both the stress cue and vowel harmony mismatch might coincide with the same word boundary. This is exactly what happens between the first two words in a sentence *Marja näkee äidin* ('Marja sees mother'). The last syllable of *Marja* contains a back harmony vowel, and the first syllable of the verb *näkee* contains a front harmony vowel. In addition, the first syllable of *näkee* is lexically stressed. Thus, word boundary is indicated by both stress and vowel harmony.

Based on behavioral evidence, Vroomen et al. (1998) and Tuomainen et al. (submitted) proposed that stress is the primary cue. Vowel harmony is only used when stress is mislocated (due to, for example, contrastive sentence accent, focus etc.) or is not realized acoustically (see Vroomen et al, 1998 Experiment 3). Based on the large difference in reaction times (RTs) between conditions as a function of the stress position, they suggested that the reason why stress is preferred for vowel harmony was that information about prominence of a syllable is available earlier than information about vowel harmony mismatch.

MODELING THE EFFECTS OF SEGMENTATION CUES IN FINNISH

The three models mentioned earlier, Cohort (e.g., Marslen-Wilson, 1984), TRACE (McClelland & Elman, 1986), and Shortlist (Norris, 1994), are all capable of lexical segmentation by providing word boundaries as by-products of the word recognition process. However, even though all these models differ in important ways in terms of their general architecture, the relevant comparison of the models for the current study concerns the issues how additional cues to word boundaries are dealt with. In this comparison, the Cohort model and TRACE will be disadvantaged over Shortlist; in their present form, Cohort and TRACE do not handle additional segmentation cues, such as the alternation of strong and weak syllables in English (e.g., Vroomen & de Gelder, 1995). Although also Shortlist in its original version lacked this property, a recent implementation of Shortlist accurately simulates the behavioral effects as well. The results suggest that English listeners' segmentations are facilitated (and in some instances penalized) by strong syllables (Norris, McQueen, & Cutler, 1995). The proposal is that MSS works pre-lexically and adds to the net activation of all those lexical candidates that are aligned with strong syllables. Thus, a segmentation cue directly affects lexical activation. To date, the only empirically supported model

that implements both the positive and negative effects of segmentation cues in English is Shortlist (Norris, McQueen, & Cutler, 1995, Norris, McQueen, Cutler, & Butterfield, 1997).

Simulations using interactive networks regarding the lexical segmentation cues in Finnish, word stress and vowel harmony, are lacking. However, Suomi et al. (1997) suggested that vowel harmony effects could be modeled in a similar manner as the Mervical Segmentation Strategy (MSS) in English. In essence, a detection of a vowel harmony mismatch could bias the activation process of the lexical network. The assumption here is that the detection of vowel harmony mismatch is pre-lexical. Regarding facilitatory effects of stress as a cue to word boundary, Vroomen et al. (1998) argued that given the fast reaction times (as compared to overall RT latencies in the experimental condition in which vowel harmony mismatch facilitated target detection), the stress effect probably stems from the pre-lexical level. Accordingly, a stressed syllable could be taken as the first syllable of the word, and lexical access could be initiated at the stressed syllable. Consequently, the activation level of all words beginning with that syllable could be boosted up, which would further increase the probability of survival in the competition process.

However, all the empirical evidence cited above is based mainly on behavioral techniques which give a measure of the total time it takes to process the stimulus and issue a response. In this respect, it is often difficult to pinpoint the functional processing stage from which the experimental effects stem. In this study we wanted to further investigate the details of the time course of the employment of simultaneous cues to the word boundary. To this end, both reaction times and event-related brain potentials (ERPs) were recorded in separate experiments. We were specifically interested in finding out whether ERPs would show a different electrophysiological pattern to word stress as compared to vowel harmony as cues to word boundaries in spoken Finnish.

ELECTROPHYSIOLOGY OF SPOKEN LANGUAGE PROCESSING AND LEXICAL SEGMENTATION

ERPs are voltage fluctuations of electric brain activity time-locked to specific events (usually to external stimuli such as spoken or written words). As such, ERPs provide millisecond accuracy of the time course of cognitive processes. Accordingly, ERPs have been used with success when the temporal aspects of cognitive processes have been the main focus of research. Several interesting findings have emerged regarding perception and comprehension of language (see Kutas and van Petten (1994), for a review).

Most of the ERP research using language stimuli has centered on the comprehension of written words typically presented in sentence context. More recently, spoken language processing has also received more attention, and differences and similarities in the ERPs between two modalities have been observed. The overall pattern of ERPs evoked by written and spoken words can be thought of as consisting of roughly three temporally consecutive sets of peaks and troughs in the ERP waveforms, which all have their distinctive electrophysiological properties and functional correlates. The early components, N1 and P2, occur within the first 200 – 300 ms post stimulus onset. These are followed by a sustained negativity starting around 300 ms and terminating at around 800 ms. Finally, a late slow positive deflection usually follows. The duration and amplitude of the late positivity depends greatly on the experimental manipulations and design.

The results of several studies suggest that modality specific processing yields differences in the early components during the first 200 – 300 ms after the onset of the stimulus, which can be seen most readily in differences in the scalp potential distribution, and also in the timing of the early components. Auditorily presented words evoke a broadly distributed N1 and a smaller P2. This complex occurs between 80 and 220 ms. N1 and P2 are considered as exogenous components, which are affected by the physical parameters, such as the (fundamental) frequency and intensity of the acoustic signal.

The following fairly slow negative deflection has been correlated with processing of phonological and semantic aspects of the spoken stimuli. All language stimuli evoke a negativity peaking around 400 ms (so-called N400). Most typically N400 is elicited by semantic anomalies both in written (Kutas & Hillyard, 1980) and spoken language (McCallum et al., 1984), although phonological manipulations have been suggested to affect the amplitude of N400 (e.g., Rugg, 1984; Praamstra et al. 1994). Visual N400 is most pronounced at the posterior scalp sites, and usually, but not always, the amplitude is larger on the right side. In contrast, N400 to spoken words is more sustained over frontal than posterior sites (Holcomb & Neville, 1990), and may be larger over the left than right hemisphere electrodes (Kutas & van Petten, 1994). However, similar overall appearance of the auditory and visual N400 as well as the fact that N400 is also evoked by semantic incongruities in (American) Sign Language (Neville, Mills, & Lawson, 1992) have been taken as an index of the workings of an amodal semantic system (Holcomb & Neville, 1990; Holcomb & Anderson, 1993). In most reports N400 has been attributed to controlled post-lexical integrative processes as opposed to an automatic process of lexical access. Even though the dichotomy between automatic and controlled processing may not be strictly accurate (Friedrich, Henik, & Tzelgov, 1991), it is generally assumed that automatic processing is fast-acting, occurs without intention and awareness, and

does not use limited capacity resources. In contrast, controlled processes are under the person's strategic control, and use limited-capacity resources (e.g., Shiffrin & Schneider, 1977). In this respect, one of the most compelling findings regarding the functional locus of N400 was reported by Brown and Hagoort (1993) using a masked priming paradigm. They showed that the N400 could be recorded only when the prime was not masked. In contrast, a significant behavioral priming effect was obtained both in the unmasked and masked conditions (see, however, Deacon et al. (2000) for a different result). This suggests that N400 is sensitive to post-lexical integration, and does not reflect automatic spreading of lexical activation.

The N400 complex also shows modality specific patterns, of which the earlier onset and longer lasting negativity are characteristic of the auditory modality. The early part of the auditory N400 was manifested in some reports as a clear and distinct negativity peaking around 200-250 ms post stimulus onset. Some researchers have related the early part to phonological processing, and results suggest that phonological and semantic effects are dissociable (Connolly and Phillips, 1994). However, there is evidence that semantic processing may have an effect already around 200 ms post stimulus onset when words are presented in sentence context (Van Petten et al., 1999). The earlier onset of auditory N400 as compared to the visual counterpart informs, first, about the temporal nature of spoken language. That is, sounds that constitute the word are exposed to the listener (almost) serially in a left-to-right manner over a period of several hundreds of milliseconds. In addition, the pattern tells us that spoken word recognition starts before the acoustic offset of the word (Marslen-Wilson, 1973; Holcomb & Neville, 1990; van Petten et al., 1999). Finally, Hagoort and Brown (2000) have suggested that the early negativity observed to spoken words reflects two related processes: First, surplus negativity is evoked by a (phonological) mismatch between the expected word form on the basis of the context (see also Connolly & Phillips, 1994), and second, the extra negativity indexes the activation (and possibly the competition) of the lexical candidates that are generated by the acoustic input.

The latter part of the ERP waveform consists of the late positive component (LPC), or Slow Wave (SW) that is usually present both when written and spoken stimuli are used. Functionally, the SW may be considered as a member of the so-called P300 family reflecting controlled and possibly post-lexical processing related to (subjective) expectancy, task relevance, decision making, or context updating (e.g., see Donchin & Coles, 1988 for a detailed argument about the interpretation of P3b as an index of context updating). Some researchers have regarded the SW as an index of the processing load demands induced by the task. For example, in a priming study Brown et al. (2000) suggested that irrespective of the task demands, participants always try to construct an integrated representation of the word pair. Thus, finding a

link between a prime and the related target is readily available. Consequently, the processing load is low and yields a more positive SW. For unrelated and neutral pairs, the link between the prime and target is harder to obtain, which is shown in a more negative SW. One should note that the SW is present most noticeably when an additional task (such as pressing the response button or counting specific targets etc.) is required. If the participants are engaged in a more natural task such as listening for understanding without additional task requirements, the slow wave is practically non-existent (for example, compare Figs. 3 and 6 in Brown et al., 2000). Finally, in paradigms in which morpho-syntactic aspects of the stimulus are manipulated, some researchers have correlated the SW (or P600) with second-pass parsing (that is, reanalysis and repair) of garden-path sentences (e.g., Hahne & Friederici, 1999; see Coulson et al., 1998 for a different view). However, all different interpretations explicitly or implicitly refer to controlled and/or post-lexical processing as underlying the late positive wave.

To the best of our knowledge, the current study is the first one focusing directly on lexical segmentation of spoken language using ERPs. A recent report by Böcker et al. (1999) investigated the ERP correlates of metrical stress in Dutch. As already mentioned earlier in the introduction, metrical stress has been shown to facilitate lexical segmentation by providing a cue about the word boundary. In the experiment by Böcker et al., bisyllabic words presented in isolation with different metrical structure (Weak-Strong (WS) or Strong-Weak (SW)) were used as stimuli. Thus, the participants were not required to segment the input. However, differences between conditions as a function of the metrical structure were observed. Specifically, WS words evoked a more pronounced negativity around 300 ms post stimulus onset of the stimulus as compared to SW words. They interpreted this negativity as an index of extra processing required by a less typical prosodic pattern (WS) in Dutch. In another study (Böcker et al., submitted) the same ERP pattern was also observed with pseudowords. In a condition in which the experimental real word stimuli were low-pass filtered the negativity was delayed by approximately 100 ms. Böcker et al. concluded that the main acoustic correlate of metrical stress is vowel color and not the other acoustic parameters such as intensity or fundamental frequency (see also Fear, Cutler, & Butterfield (1995) for a similar account of metrical stress in English).

The overall design of the current study is as follows. In Experiments 1A and 1B behavioral data will be obtained on how word stress and vowel harmony are used by Finnish listeners. The task used was a modification of the word spotting task (McQueen, 1996). Both real and pseudoword targets were employed. By using pseudoword targets, we hoped to find out whether there were differences between real word and pseudoword targets in the segmentation of the phonological strings. A

difference would suggest an involvement of lexical factors in the segmentation process. This might further help in setting up hypotheses about the temporal dynamics of lexical segmentation which are tested using ERPs in Experiments 2A and 2B.

EXPERIMENT 1A

In this experiment, the first syllable of the pseudoword carrier was stressed (or the most prominent syllable) (for example, /PU.katu/ *katu* is 'street', or /PU.vatu/ *vatu* is a pseudoword; stressed syllables are indicated with capital letters in the following examples; syllable boundary is indicated with a dot). Consequently, prominence provides conflicting information about the word boundary, because lexical stress in Finnish always lands on the first syllable of the word. The design of the experiment is similar to Suomi et al. (1997; Experiments 1 and 4) and Vroomen et al. (1998; Experiment 1), except that in the current experiment button press responses were also required for pseudowords. One of the interests in the current study was thus whether we could also obtain a vowel harmony effect for the pseudoword targets. If, however, an effect was only found for the real words that would suggest that lexical factors are involved in the detection of a vowel harmony mismatch.

METHODS

Participants. Ten students of Psychology in the University of Turku received course credit for participation. All were native speakers of Finnish, and reported no neurological or hearing problems.

Stimuli. Eighty monomorphemic CVCV nouns and adjectives of varying frequency (range 0.2 – 1590 / 1 million) were selected from a massive corpus (22.7 million word tokens) of written contemporary Finnish (Laine & Virtanen, 1998), and used as real word targets. Eighty pseudowords derived from these real word targets by changing one or two phonemes mainly at the beginning of the real word were used as pseudoword targets. In all experiments, the targets were presented in a context of either a harmonious (/pu.katu/) or disharmonious (/py.katu/) prefix so that no other real word was embedded in the nonsense string. The same harmonious and disharmonious prefixes were used for pseudowords (e.g., /pu.vatu/ or /py.vatu/). (See Appendix 3 for a list of the stimuli used in all experiments.)

Stimuli were recorded by one of the authors (JT) in a sound treated booth on a DAT tape. The stimuli were pronounced so that the word stress landed on the first syllable of the nonsense string (i.e., PU.katu, PY.katu) with natural Finnish intonation. The

stimuli were digitized at 22050 Hz (16 bit resolution) and saved to a hard disk in separate files. The total duration of the CVCVCV pseudoword carrier and onset of the first syllable of all targets were measured from the acoustic waveform. These measures were later used to correct the RTs for statistical analyses. However, even though the possible effects of durational differences between stimuli on RTs can be mostly sorted out by varying the point at which RTs are measured, the same type of analysis techniques may not prove reliable for electrophysiological data. In effect, the traditional way to analyze ERPs is to time-lock them to the onset of the stimulus (in the current study, to the onset of the CVCVCV carrier item). For this reason, and to anticipate ERP analyses, we tested the durations of targets in different conditions by using separate *t*-tests. The mean values of the total duration of the carrier items and targets for all experimental stimuli are presented in Table 1. The acoustic analyses indicated that the total duration of the CVCVCV carrier items containing real word targets with harmonious prefixes were 7 ms longer than carriers with disharmonious prefixes. This difference was significant ($t(158) = 2.471$, $p = .015$). As is typical of acoustic data due to a highly correlated pattern, very small differences between variables may turn out significant in statistical analyses. However, given that the difference is only 7 ms, it has no practical consequence, because the just noticeable difference (JND) for duration differences is higher (about 25% of the duration of the stimulus) in highly controlled conditions (see e.g., van Heuven & van den Broecke (1979) for JNDs of rise time temporal resolution discrimination). Furthermore, a 7 ms difference will not have any significant effects on ERP patterns between different conditions because the temporal resolution in the EEG recording is around 5 ms due to the sampling rate employed in most studies (i.e. 200-250 Hz). The corresponding values for the CVCVCV carrier items containing harmonious and disharmonious pseudowords were 510 (s.d. 23) ms and 509 (20) ms, respectively. This difference was not significant ($t < 1$).

The measurements of the durations of the targets indicated that real word targets preceded by harmonious prefixes were 6 ms longer than targets preceded by disharmonious prefixes. The pseudoword targets with harmonious prefixes were 4 ms longer than targets with disharmonious prefixes. Neither of the differences was significant ($t(158) = 1.580$, $p = .1162$; $t(158) = 1.122$, $p = .2636$).

Design and procedure. Two lists were constructed so that one participant heard each embedded (real word or pseudoword) target item only once. The type of prefix (harmonious or disharmonious) was counterbalanced across the lists. The position of each member of an experimental item pair was the same in the two lists. A short practice session of 24 trials preceded the experiment. The participants were tested individually in a quiet room. All items were presented over two loudspeakers (separated by a distance of 1 meter) located approximately 1.5 meters in front of the

participant. The inter-trial interval was 4.5 s. Participants were instructed that they would hear a nonsense stimulus, which sometimes contained a terminally embedded real word. They were asked to press the right-hand button (marked with the label "YES") as soon as they heard a real word, and then to say the word aloud. The vocal responses were checked by the experimenter to determine whether the intended word was correctly detected. If they did not hear a real word at end, they were instructed to press the left-hand button (marked with the label "NO") as fast as possible. No vocal response to pseudowords was required.

TABLE 1. Mean durations (ms) and standard deviations (in parentheses) of the carrier items and embedded targets for Experiments 1A (top) and Experiment 1B (bottom).

EXPERIMENT 1A (Stress on the 1 st syllable of the CVCVCV carrier item)				
	<i>Target type</i>			
	Real words		Pseudowords	
	harmonious	disharmonious	harmonious	disharmonious
Total	505 (20)	497 (22)	510 (23)	509 (20)
Target	388 (23)	382 (25)	390 (21)	386 (21)

EXPERIMENT 1B (Stress on the 2 nd syllable of the CVCVCV carrier item)				
	<i>Target type</i>			
	Real words		Pseudowords	
	harmonious	disharmonious	harmonious	disharmonious
Total	500 (21)	505 (21)	505 (21)	502 (21)
Target	405 (24)	409 (25)	403 (27)	401 (25)

RESULTS AND DISCUSSION

Unless otherwise stated, all analyses were conducted exactly in the same way in Experiments 1A and 1B. Even though it is not a standard practice in psycholinguistic research, we decided to measure reaction times (RT) from the onset

of the CVCVCV carrier item.² The major reason for this was to provide a basis for comparison of RTs and electrophysiological measures (see Experiments 2A and 2B). This decision reflects the typical procedure to analyze ERPs; single EEG sweeps (or epochs) are usually time-locked to the onset of the (spoken) stimulus (in this case, to the onset of the CVCVCV carrier item).

The vocal responses that did not correspond to the intended word (0%) and the outlying responses (3.6 %) were treated as errors and discarded from the RT analyses. The outlying responses were defined as RTs slower (or faster) than 2.5 standard deviation from individual means. Inspection of individual items and participants showed that the participants made more than 50 % errors on eight items (six real words, *kulu* 'expense', *halu* 'desire', *havu* 'fir twig', *humu* 'whirl', *supi* 'raccoon', *valu* 'casting' and two pseudowords, *halo*, and *kopu*). Those items were discarded from further analyses. No participant made more than 50% errors so that no participants were excluded from the study.

Analyses of Variance (ANOVAs) were performed with participants (*F1*) and items (*F2*) as random factors. In the by-participant analysis, Target type (real or pseudoword) and Prefix type (harmonious or disharmonious) were within-subjects variables, and in the by-item analysis, Target type was a between-items factor, and prefix type was a within-items factor.

The mean RTs and error rates are presented in Table 1 (upper panel). Inspection of the results reveals that disharmonious targets were detected 84 ms faster than harmonious targets ($F(1,9) = 47.88$, $p < .0001$; $F(1,150) = 14.00$, $p < .0001$). Furthermore, real word targets with a disharmonious prefix (e.g., "katu" in /PY.katu/) were detected 92 ms faster than targets with a harmonious prefix ("katu" in /PU.katu/). The corresponding value for the pseudoword targets was 77 ms. Planned comparisons indicated that a harmony effect was present both for the real ($t(9) = 4.421$, $p = .002$; $t(77) = 2.996$, $p = .004$) and pseudoword targets ($t(9) = 4.750$, $p = .001$; $t(77) = 2.996$, $p = .004$). Prefix type did not interact with Target type. A main effect of Target type was also present because the real word targets were detected 220 ms faster than pseudoword targets ($F(1,9) = 25.36$, $p < .001$; $F(1,150) = 79.42$, $p < .0001$).

² RTs were also measured from the onset and offset of the target items. Both of these procedures are widely used in psycholinguistic spoken word research. The rationale behind this approach is that spoken stimuli vary in duration, and as a consequence, some RT effects may be related to differences in the duration of the stimuli. For example, RTs are usually shorter for longer stimuli, because participants have more time to prepare for the response. In all significant aspects, statistical analyses yielded similar results to analyses performed on RTs measured from onset of the CVCVCV carrier. The RTs and error rates obtained from these measurements for both Experiment 1A and 1B are reported in Appendix 4.

TABLE 2. Reaction times (RTs) and error rates (in parentheses) in Experiment 1A and Experiment 1B to real word and pseudoword targets. RTs were measured from the onset of the CVCVCV carrier item. In Experiment 1A the first syllable of the CVCVCV carrier was stressed (e.g. /'PU.katu/ *katu* means 'street'), and in Experiment 1B the second syllable was stressed (e.g., /'pu.KAtu/). In both experiments targets items (real word and pseudoword) were preceded by either a harmonious or disharmonious prefix (/pu.katu/ vs. /py.katu/ or /pu.vatu/ vs. /py.vatu/, *vatu* is a pseudoword in Finnish).

EXPERIMENT 1A (Stress on the 1st syllable of the CVCVCV carrier item)

<i>Prefix type</i>	<i>Target type</i>	
	Real words	Pseudowords
Harmonious	1273 (26%)	1486 (15%)
Disharmonious	1182 (21%)	1409 (9%)

EXPERIMENT 1B (Stress on the 2nd syllable of the CVCVCV carrier item)

<i>Prefix type</i>	<i>Target type</i>	
	Real words	Pseudowords
Harmonious	868 (13%)	940 (4%)
Disharmonious	863 (10%)	938 (7%)

ANOVAs performed on the error rates showed a main effect of Prefix type ($F(1,9) = 13.83$, $p = .005$; $F(2,150) = 5.32$, $p = .022$) indicating that the disharmonious targets were easier to detect. Thus, there were no signs of speed-accuracy trade-off. A main effect of Target type was present but it was significant only in the by-item analysis ($F(1,9) = 3.42$, $p = .098$; $F(2,150) = 14.20$, $p < .0001$). No interaction between Prefix type and Target type was observed.

The results are in most part clear and straightforward. A vowel harmony effect was obtained both for the real and the pseudoword targets. The disharmonious targets were responded to faster and more accurately than the harmonious targets. Response speed and accuracy showed no interaction of target type with the prefix type.

Regarding the real words, the results are similar to Suomi et al. (1997, Experiment 1) and Vroomen et al. (1998, Experiment 1) although in the current experiment the size of the effect (92 ms) is somewhat smaller. Suomi et al. obtained an effect of

about 200 ms, and Vroomen et al. about 160 ms. Two differences might explain this discrepancy. First, we used a modification of the original word spotting task (McQueen, 1996). The original task is a Go-NoGo task but in our experiment the participants responded to each stimulus. In essence, our task is a traditional lexical decision task but with embedded targets. This may have reduced the size of the effect because the participants were required to respond on every trial. Second, in our experiment most of the targets were items belonging to the back harmony class. Only about 29% of the stimuli were front harmony items. In Suomi et al. and Vroomen et al. 50% of the targets were front harmony items. In both studies, a more pronounced difference between disharmonious and harmonious targets was obtained for front harmony items. The effect size for the back harmony real words in Suomi et al. was 103 ms (measured from the offset of target) and in Vroomen et al. the difference was 91 ms (measured from the offset of the target). In our experiment the difference is 77 ms when the RTs were measured similarly, which is fairly close to the difference obtained by Vroomen et al. and Suomi et al. (see Appendix 4 for detailed results). The unbalanced proportion of the front harmonious and back harmonious stimuli in the current study, however, reflects better the relative frequencies of front and back vowels in Finnish (e.g., Karlsson, 1982), and in that sense also the reality of how often words with front or back harmonious vowels will be encountered by Finnish listeners.

A new observation of the current experiment is that a similar effect of vowel harmony was also observed for the pseudoword targets. The size of the effect was 77 ms (measured from the onset of the CVCVCV carrier item), which is somewhat smaller than the one obtained for the real words. Because there was no interaction between the target type, this suggests that for most part similar processes underlie the vowel harmony effect for both the real and the pseudoword targets. Given the slow overall RTs we suggest, as did Vroomen et al. (1998), that post-lexical processing is involved that may account for the harmony effect.

EXPERIMENT 1B

This experiment differs from the previous experiment in that the position of the stressed syllable is changed. The first syllable of the target is now the most prominent syllable of the nonsense carrier item (for example, /pu.KAtu/ or /py.KAtu/). The present experiment thus closely resembles the second experiment of Vroomen et al. (1998) except that in the present experiment a speeded response is also required to pseudowords (similar to Experiment 1B of the current study). The main reason for this modification was the same: comparison between RTs to

pseudoword and real word targets might provide information about the possible time course of how word stress information and vowel harmony mismatch are exploited by listeners in detecting word boundaries in connected speech. For example, if the correct stress position would speed up responses in real words only but not in pseudowords, that would indicate that stress is possibly coded lexically, or that other processing requiring lexical information is responsible for the stress effect. In contrast, a similar stress effect for the real and the pseudoword targets would suggest a common functional locus of the computation of the prominence of the syllable.

METHODS

Participants. Ten students different from those attending the preceding experiment participated. All were students of Psychology at the University of Turku and they received course credit for participation. All were native speakers of Finnish, and reported no neurological or hearing problems.

Stimuli. The same stimuli as in the preceding experiment were used. However, the stimuli were re-recorded (spoken by JT) in the same recording session as the stimuli of the preceding experiment. Acoustic analyses indicated that the total duration of the CVCVCV carrier items containing real word targets preceded by harmonious prefixes was 5 ms longer than targets preceded by disharmonious prefixes (see bottom part of Table 1 for duration values). For the carrier items containing pseudoword targets a 3 ms difference was observed. Comparison of the durations showed no significant differences ($t(158) = -1.553$, $p = .122$; $t(158) = .914$, $p = .362$, respectively).

Additional measurements of the duration of the target items revealed that real word targets preceded by harmonious prefixes were 4 ms shorter than targets preceded by disharmonious prefixes ($t(158) = 1.032$, $p = .3035$). Pseudoword targets with harmonious prefixes were 2 ms longer than targets with disharmonious prefixes, and this difference was not significant ($t < 1$).

Since we are essentially interested in temporal factors, and between experiment comparisons are of major importance, durational differences between the stimuli used in the two experiments might account for (some of) the effects, especially in the electrophysiological recordings. Accordingly, a 2x2x2 ANOVA with Target type as between-item factor and Stress position and Prefix type as within-item factors was performed separately for total and target duration values. As expected, the 2 ms difference in total duration between the stimuli of Experiment 1A and 1B showed only a trend (Main effect of Stress ($F(1,316) = 3.061$, $p = .082$). Stress position was involved in several interactions that were significant, but because in none of them the

durational differences were larger than 13 ms, they will be ignored as functionally not significant.

A similar ANOVA performed on the duration values of the target items indicated no significant difference in the durations between the real and the pseudoword targets ($F < 1$). However, targets with stressed first syllable were 18 ms longer than unstressed targets, which yielded a significant main effect of Stress position ($F(1,158) = 172.329$, $p < .0001$). More importantly, the second-order interaction of Stress by Prefix type by Wordtype was not significant ($F(1,158) = 2.558$, $p = .112$), indicating that durational differences between stimuli used in these two experiments were small.

Design and procedure. The design and the procedure were exactly the same as in Experiment 1A.

RESULTS AND DISCUSSION

The reaction times were measured in the same way as in Experiment 1A, that is, from the onset of the CVCVCV carrier item (see Appendix 2 for results for RTs measured from the onset and offset of the target item). RTs faster or shorter than 2.5 s.d. from individual means (4.3 %) were excluded from further analysis. An examination of the individual results revealed that none of the participants made more than 50% errors, and thus none were excluded. However, four items (all containing real word targets; *kulu* 'expense', *salo* 'backwoods', *kulo* 'forest fire', *supi* 'raccoon') were excluded because over 50% of the participants made an error in these items.

The RTs and error rates are presented in the bottom panel of Table 1. Inspection of the results shows that disharmonious real word targets were detected only 5 ms faster than harmonious targets. Similarly, disharmonious pseudoword targets were detected 2 ms faster. Accordingly, neither a main effect of Prefix type nor an interaction with Target type was observed (all F 's < 1). Real word targets were responded 84 ms faster to than pseudoword targets ($F(1,9) = 21.16$, $p < .001$; $F(1,156) = 35.71$, $p < .0001$).

ANOVAs on the error rates showed no main effect of Prefix type (both F 's < 1). As with the reaction times, a main effect of Target type ($F(1,9) = 10.60$, $p = .010$; $F(1,156) = 5.67$, $p = .019$) was present. However, in this case, the participants made more errors on real words. Thus, for some reason, a speed accuracy trade-off might be present regarding the real and pseudoword targets. Furthermore, the participants responded more accurately to the disharmonious real word targets (3%), but in the pseudowords the pattern was reversed (3% more accurate on harmonious targets). The interaction approached significance in the by-item analysis ($F(1,9) = 2.94$, $p = .121$; $F(1,154) = 3.48$, $p = .064$).

These results show convincingly that no vowel harmony effect was obtained in this experiment. The pattern of data is very similar to Vroomen et al. (1998; Experiment 2) and Tuomainen et al. (submitted). The common denominator in all of these experiments is that the first syllable of the target was the most prominent syllable of the carrier item. As a consequence, prominence provided correct information about the onset of the target, which again suggests that listeners prefer word stress to vowel harmony as a cue to word boundary. What is notable in the results is that a similar effect was obtained for the pseudowords, suggesting that the computation of stress is (at least partly) a pre-lexical process. However, before these conclusions can be confirmed in the current study, we decided to compare the results between Experiment 1A and 1B. This will provide us with a direct test on how word stress and vowel harmony behave as a function of target type and location of the word stress.

COMPARISON OF THE RESULTS BETWEEN EXPERIMENT 1A AND 1B

The reaction times were measured from the onset of the CVCVCV carrier item³. A 3-way ANOVA was performed separately with participants ($F1$) and items ($F2$) as random factors. In the by-participant analysis the Stress position (first vs. second syllable of the CVCVCV carrier item) was a between-subject factor and Target type (real word vs. pseudoword) and Prefix type (harmonious vs. disharmonious) were within-subject factors. In the by-item analysis, Target type was a between-item, and Stress position and Prefix type were within-item factors. To make the database comparable for statistical analyses, the same 10 items were excluded from each experiment because of excessive errors. (It should be noted that of the total of 12 items excluded in both experiments in separate analyses, two items, *kulu*, 'expense', and *supi* 'raccoon', were discarded in both experiments).

Inspection of Table 1 shows that the overall RTs in Experiment 1B were over 400 ms faster as compared to Experiment 1A, which resulted in the main effect of Stress position ($F1(1,18) = 44.57, p < .0001$; $F2(1,148) = 56.32, p < .0001$). Moreover, main effects of Target type (real word targets were recognized 146 ms faster than pseudoword targets; $F1(1,18) = 40.44, p < .0001$; $F2(1,148) = 13.41, p < .0001$), and Prefix type were obtained (disharmonious targets were detected, on the average, 43 ms faster than harmonious targets ($F1(1,18) = 45.51, p < .0001$; $F2(1,148) = 14.31, p = .0001$)).

Two interactions were also observed. First, and most importantly, Stress position interacted with Prefix type ($F1(1,18) = 38.26, p < .0001$; $F2(1,148) = 11.94, p = .001$).

³ Between experiment comparisons were also performed on the RTs measured from the onset and offset of the target item. In all significant aspects, the results did not differ from those reported in the main text.

This confirms the earlier finding that vowel harmony effect was obtained only when prominence provided incorrect information about the word boundary (that is, in Experiment 1A in which the first syllable of CVCVCV carrier item was stressed). This result is also exactly the same as obtained by Vroomen et al. (1998). However, vowel harmony effect did not interact with Target type. In contrast, Stress position had a more pronounced effect on the detection of the pseudoword targets than to the real word targets ($F(1,18) = 9.74$, $p = .006$; $F(1,148) = 35.00$, $p < .0001$). The RTs to the pseudowords were, in general, 508 ms faster in Experiment 1B as compared to 1A. The corresponding value for the real word targets was 362 ms. Thus, the pseudowords profited more from correct stress than did the real words. No second-order interactions were observed.

ANOVAs performed on the error rates showed the following main effects. First, more errors were made in Experiment 1A than in Experiment 1B ($F(1,18) = 14.14$, $p < .001$; $F(1,148) = 56.32$, $p < .0001$). Second, the targets with harmonious prefixes were on the average more difficult to detect than targets with disharmonious prefixes ($F(1,18) = 5.44$, $p = .031$; $F(1,148) = 5.00$, $p = .027$). Third, more errors were made when the real words were targets ($F(1,18) = 6.81$, $p = .018$; $F(1,148) = 13.41$, $p < .0001$).

Similar to the RTs, ANOVA on error rates yielded an interaction between Stress position and Prefix type ($F(1,18) = 4.59$, $p = .046$; $F(1,148) = 5.18$, $p = 0.24$). In contrast to the RTs, Stress position by Target type interaction was (marginally) significant only in the by-item analysis ($F(1,18) = .83$, $p = .374$; $F(1,148) = 3.88$, $p = .051$). This trend relates to the fact that when stress provided correct information about the word boundary, accuracy increased more in the real words than in the pseudowords.

GENERAL DISCUSSION OF THE BEHAVIORAL RESULTS

Two reaction time experiments were performed in which the effect of word stress and vowel harmony as cues to word boundary in Finnish was investigated. The results showed that word stress seems to be the primary cue, and only when the stress cue is not available, listeners resort to vowel harmony. These results are identical to the ones obtained by Vroomen et al. (1998) and Tuomainen et al. (submitted). Thus we have now evidence from three different sources (and using two different methodologies, see Vroomen et al., 1998, Experiment 3, for using synthetic speech with prosodic and phonotactic alterations) that all point to the same conclusion. Furthermore, in the current experiments, we showed that with the pseudoword targets a similar pattern to the real words was obtained in terms of how stress and vowel harmony are employed by the listeners. In essence, reaction times were speeded up for both the real words and pseudowords, and more so for the latter. Our conclusion is that the computation of

prominence is fast and stress information is early available to the listener. Furthermore, because no significant differences were observed between the real word and pseudoword targets when stress provided a correct cue about the word boundary, all explanations related to the lexical status of the targets (and hence any usage of information related to the lexical entry) can be discarded. This finding might be taken as an indication that the prominence level of a syllable is detected pre-lexically.

The reliable vowel harmony effect seems to appear only when the participants respond slowly. The average response time from the offset of the stimulus for the real word targets is about 700 ms, and for the pseudoword targets about 950 ms (see Appendix 4). This is about the same time interval that is required for a response measured from the onset of a (three-syllabic) spoken word in a lexical decision task.

With behavioral techniques we have shown that the detection of a word boundary in connected speech is facilitated by word stress and vowel harmony mismatch. Based on reaction times, we suggest that word stress is the primary cue simply because it can be computed faster, and is thus earlier available to a listener. If, however, word stress provides conflicting information about a potential boundary, then listeners rely on vowel harmony. The detection of the harmony mismatch seems to be a time consuming process, which suggests that controlled and strategic processes may be involved.

We will now turn to a different methodology, namely event-related potentials (ERPs) which have one main advantage over reaction times. With ERPs, we have a window (even though an indirect one) to cognitive processes with a millisecond precision. Reaction times measures inform us about the end product of the cognitive process including stages of perception, evaluation and issuing a response. In contrast, ERPs provide us information about the time course of brain processes as they take place in real time. However, we consider these methodologies complementary. The main goals of the following two experiments are to find converging evidence for the behavioral results of the earlier experiments, and to explore in more detail the time course of how multiple cues to word boundaries are utilized during the recognition of words in continuous speech. The question is whether prominence and vowel harmony as cues to word boundaries have different consequences, which show up in the ERPs. If the perception of stress takes place early (and possibly) pre-lexically, and the detection of vowel harmony mismatch is a long-lasting process, the ERPs should show stress effects earlier than vowel harmony, and mostly before the acoustic offset of the spoken stimulus.

EXPERIMENT 2A

In this experiment the same stimuli were used as in Experiment 1A. The task of the participants was again to try to recognize the end-embedded target stimuli. While they were performing the task, voltage fluctuations of the brain related to word recognition were recorded by using scalp electrodes. The original instruction to respond as fast as possible was, however, changed to a delayed response task. The idea was to exclude possible artifacts created by brain activity related to response preparation and motor potentials caused by the button press. This change seems especially warranted since the response times in Experiments 1A and 1B as a function of stress position were extremely different. Consequently, if we had measured RTs at the same time with ERPs, it might have had a drastic effect on the overall architecture of the ERPs because of anticipatory and motor potentials. This would have made the comparison of waveforms between the experiments difficult and possibly unwarranted.

METHODS

Participants. Eleven students (2 males) from the introductory course of Psychology participated for course credit in the experiment. All were native speakers of Finnish with no known neurological or hearing disorders. All except one were right-handed. After preliminary data analysis, one participant was excluded because of excessive background alpha activity during the recording, thus leaving us with a total of 10 participants.

Stimuli. The same stimuli used in Experiment 1A were employed in this experiment.

Procedure. The subjects were run in a dimly lit EEG cabin in a single session that lasted about 1.5 hours (the duration of the experimental session was about 1 h). Prior to the experiment, the participants were given an instruction sheet describing the course of the experiment. The experiment consisted of one practice block and one experimental block. A total of three short pauses (one after a presentation of 40 trials) were held during the experimental block. One trial consisted of three events: first a warning signal (a red circle) was displayed for 500 ms in the center of the computer screen. After a variable SOA (ranging from 750 ms to 1250 ms) the experimental stimulus (a spoken pseudoword) was presented through loudspeakers positioned approximately 1.5 meters in front of the participants. Two seconds after the offset of the auditory stimulus, a “GO”-signal (a green circle) was displayed for 2 seconds, during which the participants were to respond. A fixation point was displayed throughout the experiment in the center of the computer screen.

The participants were told that they would hear nonsense "words" that occasionally contained a real word embedded at the end of the nonsense item. The participants were asked to wait until a "GO"-signal appeared on the computer screen in front of them, and to press with their index finger a response button (labelled "YES") if they had heard a real word at the end of the pseudoword. After the button press they were required to say aloud the word that they had heard. The oral responses were noted down by the examiner. If, however, they did not hear a real word, they were instructed to press the left button (labeled "NO"). A 2 sec pause preceded the next trial. There was a short practice block of 24 items.

The participants were also instructed to keep their eyes focused on the fixation point during the time the red warning signal was on. They were further asked to restrain from moving their eyes, blinking, swallowing etc. during the time interval starting from the warning signal and ending to the "GO" signal. Blinking between trials was encouraged. The response hand was counterbalanced across subjects.

The ERP recording. The electroencephalogram (EEG) was recorded from 15 Ac/AcCl scalp electrodes referenced to linked mastoids, which also served as ground. Nine electrodes were placed on standard International 10-20 System locations - left and right parietal (P3, P4), left and right central (C3, C4), left and right frontal (F3, F4), and midline (Fz, Cz, Pz). Six electrodes were placed on nonstandard locations to cover traditional language areas. These sites include Wernicke's region and the right hemisphere homologue (WL, WR), left and right temporal (TL, TR) and anterior temporal left and right (ATL, ATR). Horizontal eye-movements were recorded by attaching one electrode on right outer canthus, and vertical eye-movements and blinks were recorded by an electrode attached below the left eye. Both EOG electrodes were referenced to linked mastoids. Electrode impedance was monitored before and after the recording, and was kept below 5 K Ω . The EEG was amplified by a Braintronics amplifier (band pass 0.5-70 Hz). Continuous EEG was digitized at 200 Hz and stored on a hard disk for later analysis.

Data analysis. Continuous EEG was epoched to 1200 ms trials containing a 100 ms baseline before stimulus onset. After the baseline correction, trials contaminated by artifacts were automatically rejected ($\pm 70 \mu\text{V}$). The trials on which participants responded incorrectly were also excluded from the analyses. This resulted in a total loss of approximately 10% of all trials. (In Experiment 2B approximately 12% of the trials were excluded for similar reasons.) For each participant, averaged ERPs were calculated separately for each of the four conditions (unstressed harmonious and disharmonious real word targets; unstressed harmonious and disharmonious pseudoword targets). The individual

averages were band-pass filtered (0.5-25 Hz, 24 dB cutoff slope) and baseline corrected after filtering. All statistical analyses were performed on the raw data. For illustrative purposes only, the grand-average ERPs were smoothed off-line using a 6 Hz low-pass filter.

The averaged waveforms were quantified by computing the mean amplitude of four consecutive 100 ms time windows starting at 200 ms after the onset of the spoken stimulus. The inspection of the waveforms indicated that the early effects related to experimental manipulation coincided with this time frame. Also, based on the acoustic measures, at 300 ms, participants had heard more or less completely the first two syllables. Before this time point, possible effects observed in the waveforms would probably reflect exogenous responses to the acoustic features (such as differences in (intrinsic) intensity, spectral quality, and (intrinsic) duration between phonemes). Thus, after 300 ms from the onset of the stimulus enough information about the phonological structure of the stimuli could have been gathered so that, for example, a detection of vowel harmony mismatch could be possible. In addition to these four successive time windows, the late positivity observed in the waveforms was quantified by computing the mean amplitude of two time windows between 700 and 900 ms and 900 and 1100 ms. Each time window was subjected to separate repeated measures Analyses of Variance (ANOVAs). The midline electrodes were not analyzed, because inspection of the waveforms indicated that brain responses at midline did not differ from the other electrodes.

Brain responses as a function of electrode position (along the anterior-posterior axis) or laterality (left vs. right) will usually show highly significant results. These main effects are not directly related to the major focus of the current study, and we will not discuss the effects of Electrode position and Laterality unless they are involved in an interaction with Target type (Word vs. Pseudoword target) and/or Prefix type (Harmonious vs. Disharmonious prefix). Interactions in ANOVA involving factors Electrode position and/or Laterality are usually taken as an indication of different underlying current sources. However, McCarthy and Wood (1985) showed that because of the incompatibility between an additive (linear) ANOVA model and the multiplicative (non-linear) effect of the changes in the electric source strength, this assumption may result in an incorrect finding. Thus, Condition by Location interaction may not reflect genuine source differences but only differences in the strength of the same source. They suggested that one way to circumvent the problem is to scale the data separately for each condition and perform the ANOVA with scaled data. Accordingly, the current data were normalized by using a z-score standardization (see also, Rösler, Pütz, Friederici, & Hahne, 1993). In effect, the differences in mean and variance between conditions were removed separately for each measurement epoch.

The normalized values reflect the standardized deviations of the amplitude from the grand mean across electrodes. The unwanted consequence of normalization is that all main effects and interactions involving Prefix type and Target type (and Stress location in the between experiment comparison) are zero. Consequently, the ANOVAs were first performed using the raw data. Additional analyses were then performed with the scaled data to investigate the scalp potential distribution regarding vowel harmony and stress effects. Greenhouse-Geisser correction of the degrees of freedom and Bonferroni corrections were used when appropriate. For the Greenhouse-Geisser correction, original degrees of freedom together with the epsilon correction factor and the corrected probability level are reported.

RESULTS

Behavior

On the average, the participants responded correctly to 97% of all trials (range 87.5 – 100%). As in the reaction time experiments, the participants were more accurate on the pseudowords (98.4%) than on the real word targets (95.4%). Also, and somewhat surprisingly, the responses to the harmonious targets (98.2%) were more accurate than to the disharmonious targets (95.7%). However, the good overall performance indicates that participants attended to the stimuli during the recording session.

ERPs to real word targets

Figure 1 shows the averaged waveforms to the harmonious and disharmonious real word targets. (In all figures, the acoustic onset and offset of the targets have been marked with vertical dotted lines so as to help relate different ERPs effects to these temporal landmarks.) Inspection of the waveforms reveals that between the onset and 200 ms both stimuli evoked similar fronto-central N1 (peaking at about 100 ms post stimulus onset) and P2 (at about 200 ms) deflections that are typical of auditory ERPs. A long lasting negativity follows. This negativity starts around 300 ms and returns back to the baseline at around 850 ms after the stimulus onset at the frontal and at around 750 ms after the stimulus onset at the posterior electrodes. It is most pronounced at frontal and central electrodes, probably, at least partly, reflecting an anticipatory (CNV type) response.

Repeated measures ANOVAs performed on the six time intervals showed two separate time frames in which significant effects related to the Prefix type were obtained. First, an early effect occurring between 400 and 500 ms (slightly before

the acoustic offset of the spoken stimuli). During this interval responses to the disharmonious targets were more negative at the left sided electrodes (Prefix type by Laterality interaction, $F(1,9) = 8.824$, $p = .016$). However, separate t-tests revealed that using a Bonferroni corrected alpha level (.007) the difference between the conditions was significant only at the left Anterior Temporal (ATL) electrode ($t(9) = 3.693$, $p = .005$). A late effect between 700 and 900 ms (that is, a few hundred milliseconds after the acoustic offset of the stimuli) resulted in a significant main effect of Prefix type ($F(1,9) = 8.019$, $p = .020$), and in an interaction between Prefix type and Electrode position ($F(6,54) = 5.546$, $p = .011$, $\epsilon = .359$). This was due to a smaller effect at the lateral (WR/WL, T4/T3, ATR/ATL) as compared to the more medial electrode sites.

ERPs to pseudoword targets

The ERPs to the pseudoword targets are presented in Figure 2. The overall pattern of the ERPs is similar to the ERPs obtained with the real word targets. A similar N1 and P2 deflection were observed both for the harmonious and disharmonious targets followed by a long lasting, but especially at the posterior electrodes, a smaller negativity. The onset of this negativity is at around 300 ms (similar to real words) but it returns back to baseline at the fronto-central electrodes between 950 and 1000 ms, reflecting longer processing as compared to the real word targets. Slightly more negative responses between 400 and 500 ms at the left anterior electrodes to the disharmonious targets can be noted, but, in contrast, responses to the harmonious targets are more negative at the left posterior electrodes. This pattern yielded a significant two-way interaction of Prefix type and Electrode position ($F(6,54) = 4.557$, $p = .019$, $\epsilon = .381$), and a three-way interaction of Prefix type, Electrode location and Laterality ($F(6,54) = 3.020$, $p = .046$, $\epsilon = .508$).

Real word vs. pseudoword targets

The results described above suggest that the harmony effect was more pronounced for the real words. To see whether the same pattern was also found when the ERPs to the real and pseudoword targets were directly compared, we conducted another ANOVA using the mean amplitude values with Target type, Prefix type, Laterality (left vs. right), and Electrode position (7 levels in the anterior-posterior axis) as factors. Separate analyses were performed using normalized data when significant interactions involving Prefix type, Electrode position and/or Laterality were observed.

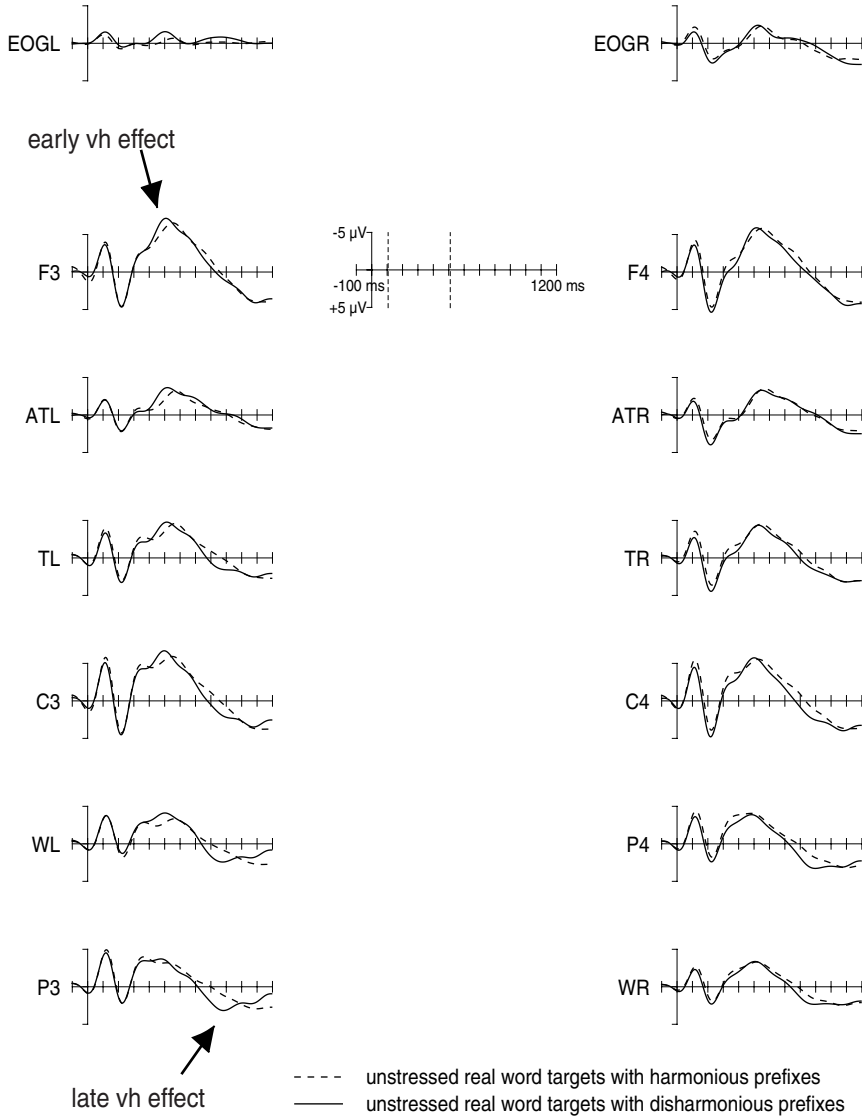


FIGURE 1. ERPs to real word targets preceded by harmonious (broken line) and disharmonious (solid line) prefixes in Experiment 2A (unstressed targets). The onset of the nonsense carrier item is at zero milliseconds. More negative responses to disharmonious targets between 400 and 500 ms after the onset of the stimulus are present at the left-sided electrode locations. The arrow at the left anterior temporal (ATL) electrode points at the position where the early effect is largest. A later effect between 700 and 900 ms (indicated by an arrow at the left parietal electrode) shows more negative responses to harmonious targets mainly at posterior electrodes, also indicated by an arrow. Vertical dotted lines mark the acoustic onset and offset of the targets. Negative polarity is plotted upwards in this and all subsequent figures.

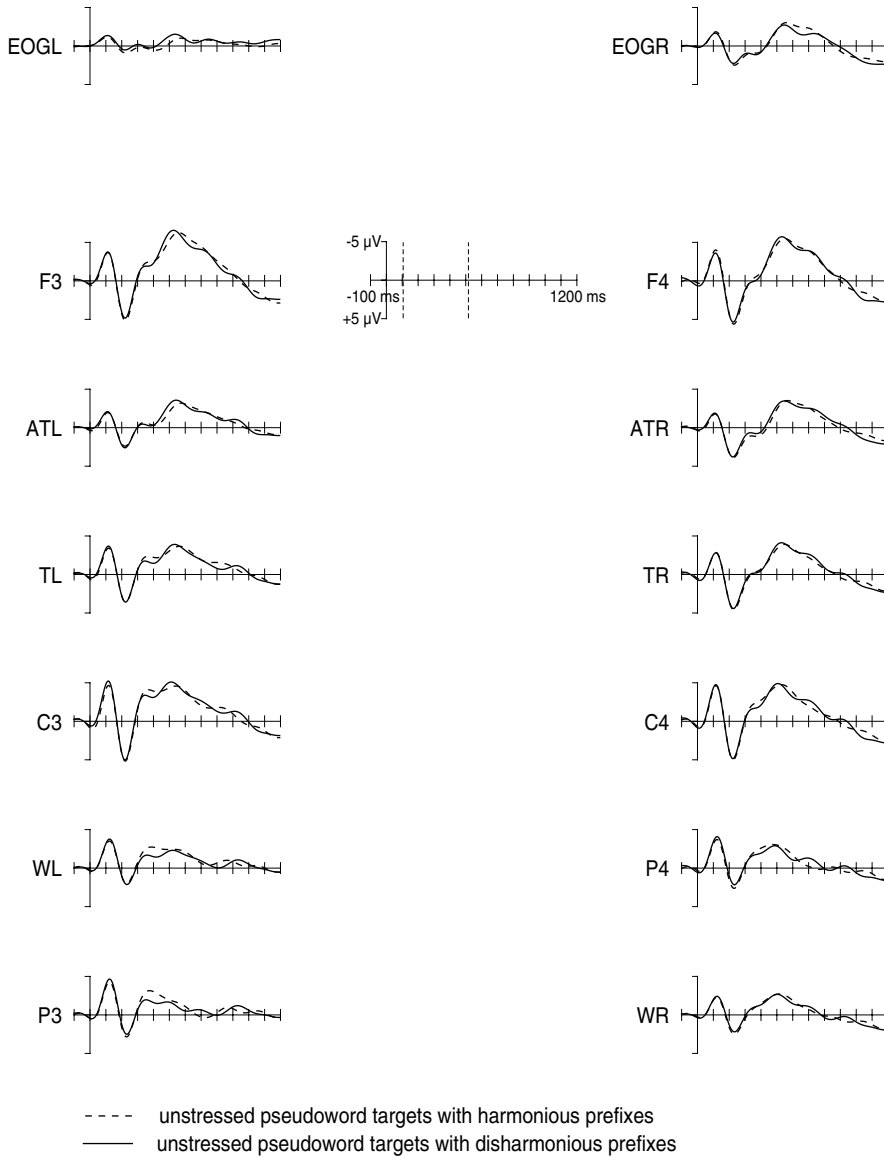


FIGURE 2. ERPs to pseudoword targets preceded by harmonious (broken line) or disharmonious (solid line) prefixes in Experiment 2A (unstressed targets). The onset of the nonsense carrier item is at zero millisecond. Responses to disharmonious targets are more negative at left frontal (F3 and ATL) electrodes between 400 and 500ms. In contrast, in the same time window, at the posterior electrodes, an opposite effect can be noted; responses are more negative to harmonious targets. No late difference between waveforms can be observed. Vertical dotted lines mark the acoustic onset and offset of the targets.

Harmony effect in the early time frame. No main effect of Prefix type was observed in the early time windows. However, between 400 and 500 ms the disharmonious targets were more negative at the anterior electrodes (Prefix type by Electrode position ($F(6, 54) = 4.996$, $p = .024$, $\epsilon = .293$). Furthermore, a complex interaction involving Prefix type, Target type and Electrode position and Laterality was observed. This pattern is depicted in Figure 3. The graph shows amplitude values from selected left sided electrode locations between 400 and 500 ms. The results showed, first, for real word targets that the disharmonious targets yielded more negative amplitudes at all electrodes. However, for the pseudoword targets, the anterior electrode shows a more negative response to targets preceded by the disharmonious prefixes but at the central and posterior electrode the harmonious targets yielded more negative responses. This pattern of data resulted first of all in a three-way interaction of Target type by Prefix type by Laterality ($F(1,9) = 5.442$, $p = .045$), and also in a (trend for) four-way interaction of Target type by Prefix type by Electrode location by Laterality ($F(6,54) = 3.006$, $p = .053$, $\epsilon = .491$).

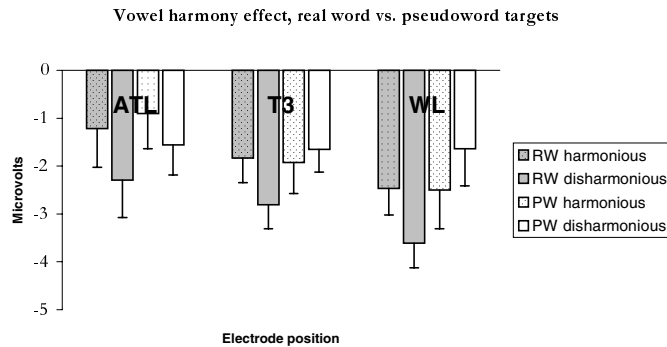


FIGURE 3. A bar graph illustrating ERP differences between real and pseudoword targets preceded by harmonious or disharmonious prefixes. Grey bars denote real word and white bars pseudoword targets. In all selected electrodes, more negative responses to real word than pseudoword targets are observed. For pseudoword targets, a cross-over interaction is noted: responses are slightly more negative to disharmonious targets at ATL, but at WL (a posterior electrode at around Wernicke's area on the left) responses are more negative to harmonious targets.

Harmony effect in the late time frame. A main effect of Prefix type was present in the late time window between 700 and 900 ms ($F(1,9) = 6.873$, $p = .028$). In the

same time window there was a trend toward more positive responses to the disharmonious targets as compared to the harmonious targets at the posterior electrodes (Prefix type * Electrode position, $F(6,54) = 3.022$, $p = .060$, $\epsilon = .411$). Also, Target type by Prefix type interaction was present which was due to a slightly larger difference between the harmonious and disharmonious real word targets ($F(1,9) = 5.058$, $p = .051$). Furthermore, inspection of the waveforms suggests that the harmony effect in this time frame is weak for the pseudoword targets but fairly robust for the real word targets especially at the posterior electrodes. However, an interaction involving Electrode position was not significant in the ANOVA.

Main effects and interaction of Target type with electrode position. A significant main effect of Target type was obtained between 300 and 500 ms (more negative responses to real word targets), and 700 and 1100 ms (more negative responses to pseudoword targets) (all p 's < .05). Target type interacted with Electrode position between 900 and 1100 ms which was due to more negative responses to the pseudowords at posterior, and especially at the parietal electrodes ($F(6,54) = 10.66$, $p < .0001$, $\epsilon = .381$).

DISCUSSION

The ERPs showed differences in the processing of the harmonious and disharmonious targets in two time frames. The first one, between 400 and 500 ms coincides with an interval slightly before the acoustic offset of the stimuli. The ERP responses to the targets with disharmonious prefixes were more negative than to the targets with harmonious prefixes. The second interval, between 700 and 900 ms, displayed a more negative response to the targets with harmonious prefixes than to the targets preceded by disharmonious targets thus suggesting that the detection of the target item in the harmonious context is more difficult.

Significant differences in the processing of the real and pseudoword targets were also observed. First, processing of the harmony mismatch in the early time frame recruited larger left-sided areas for the real words. For the pseudowords, only weak evidence for the vowel harmony effect was observed. Second, later effects were present only for the real words most notably at the posterior electrodes. These observations suggest that the processing of the harmony mismatch begins soon after the mismatching vowel has been heard. Furthermore, the difference in the vowel harmony effect between the real words and the pseudowords could be accounted by assuming that, for the real words, the detection of the vowel harmony mismatch may involve the activation of the phonological code stored in the form based lexical representation.

Finally, before moving on to the final experiment, one should keep in mind that the reaction times obtained in Experiment 1A were extremely slow both for the real words (about 1100 ms) and for the pseudowords (about 1400 ms) (see Table 1). This finding does not clearly show up in this experiment. What we see here is that especially the real word targets preceded by disharmonious prefixes yield more negative responses fairly early, as soon as enough information about the vowel qualities are available. This negativity probably signifies additional processing that might be related to the detection of the harmony mismatch. A reverse pattern is observed in the late time frame; more negative responses to the harmonious targets are observed, this time suggesting that more processing is required for segmenting targets with harmonious prefixes. Thus, the disharmonious targets have a processing advantage over the harmonious ones at this time interval. One cannot, however, conclude that the disharmonious targets have already been recognized at this point. It is possible that the processing of both the harmonious and disharmonious unstressed targets is still going on. We lack the correct control condition, which could verify this assumption. As we will see later on, a comparison to the stressed real word targets indicates that the processing load of the unstressed targets preceded by disharmonious prefixes is higher in the late window. Thus, before discussing further the results of Experiment 2A we turn to Experiment 2B in which ERPs to stressed target stimuli (i.e., the stress is located on the second syllable of the carrier item) were recorded.

EXPERIMENT 2B

The purpose of the final experiment is to investigate the temporal pattern of the brain responses evoked by target stimuli in which the first syllable of the target was stressed. The reaction time data obtained in Experiment 1B indicated over 400 ms faster responses to these types of stimuli, and no harmony effect was present for either real or pseudoword targets. Thus, based on the results of the preceding experiments, we would not expect to find early differences between harmonious and disharmonious targets, but only an effect of target type that would reflect the fact that the real word targets were detected faster than the pseudoword targets.

METHODS

Participants. Another 10 students (2 males) from the introductory course of Psychology participated for course credit in the experiment. All were right-handed and native speakers of Finnish with no known neurological or hearing disorders.

Stimuli. The same stimuli used in Experiment 1B were employed in this experiment.

Procedure. The procedure was exactly the same as in Experiment 2A.

ERP recording and Data analysis. Exactly the same recording technique and data analysis procedure as in Experiment 2A were used.

RESULTS AND DISCUSSION

Behavior

Also in this experiment the participants correctly responded to 97% of all stimuli (range 85 – 100%). Again as in Experiment 2A and in the reaction time experiments, participants were more accurate on pseudowords (99.7%) than on real word targets (94%), and similar to the preceding experiment, the responses to the harmonious targets (98%) were more accurate than to the disharmonious targets (95.8%). The overall error rates are identical between the experiments showing that in both experiments the participants attended to the stimuli during the recording session.

ERPs to real word targets

Figure 4 shows the averaged waveforms to the harmonious and disharmonious real word targets. Inspection of the waveforms reveals the same overall pattern as in Experiment 2A; between the onset and 200 ms both stimuli evoked similar fronto-central N1 (peaking at about 100 ms post stimulus onset) and P2 (at about 200 ms) deflections that are typical of auditory ERPs. These deflections were followed by a long lasting negative wave, which begins to deviate from the baseline at around 300 ms at all electrode sites. It returns to baseline at frontal electrodes at around 800 ms and at posterior electrodes at 800 ms. A late positivity can be seen most clearly at the posterior electrodes. The duration of this positivity is approximately 500 ms. The late positivity peaks at the parietal electrodes at around 820 ms. Figure 4 reveals that between 300 and 500 ms the disharmonious targets show a more negative response as compared to the harmonious targets especially at some of the lateral and posterior electrodes mostly over the left hemisphere (T3, WL, P3, and P4). However, repeated measures ANOVA with Prefix type, Electrode position (7 levels) and Laterality revealed no significant main effects or interactions involving Prefix type (all F 's < 1, except between 300-400 ms, Prefix

by Electrode position interaction, $F(6,54) = 2.696$, $p = .093$; Prefix by laterality by Electrode interaction, $F(6,54) = 2.260$, $p = .103$; and between 400-500 ms, Prefix by Electrode by Laterality interaction, $F(6,54) = 1.982$, $p = .152$). This indicates that there were no differences in the processing of the real word targets with harmonious and disharmonious prefixes.

ERPs to pseudoword targets

ERPs to the pseudoword targets are presented in Figure 5. Once again, similar N1 and P2 deflections were observed for both targets preceded by harmonious and disharmonious prefixes. A negativity starting at 300 and returning back to baseline at around 800ms at the fronto-central electrodes, and at around 750 ms at the posterior electrodes. The peak amplitude is highest at the posterior electrodes around 900 ms with a less well defined peak for the disharmonious targets.

The responses to the harmonious targets are slightly (but not significantly) more negative between 300 and 500ms (all F 's < 1). A reverse pattern is observed between 600 and 700 ms, and this difference is more pronounced at the left sided electrodes (Prefix type by Laterality interaction, $F(1,9) = 5.967$, $p = .037$). No other significant differences were noted.

Real word vs. pseudoword targets

As in Experiment 2A, we compared the responses to real word and pseudoword targets by running a repeated measures ANOVA. The main interest was to find out whether there were any differences as a function of Target type. Normalized amplitude values were employed to test the differences in the scalp distribution.

Main effects and interactions involving Target type. Significant differences between Target type were noted in the late time windows between 600 and 900 ms (all $p < .05$). This was due to more negative responses to the pseudoword targets. In the last time window (900 to 1100 ms) the difference approached significance ($F(1,9) = 4.964$, $p = .054$). Target type interacted with Electrode position in two separate time frames, first, between 500 and 600 ms more negative responses to the pseudowords were recorded at frontal electrodes, which approached significance ($F(5, 64) = 3.769$, $p = .054$, $\epsilon = .372$), and second, between 700 and 1100, responses to the pseudoword targets were more negative at posterior electrodes (all p 's $< .05$). Finally, between 900 and 1100 ms the difference between the ERPs to the pseudoword and real word targets was larger at the right hemisphere electrodes ($F(1,9) = 5.944$, $p = .037$).

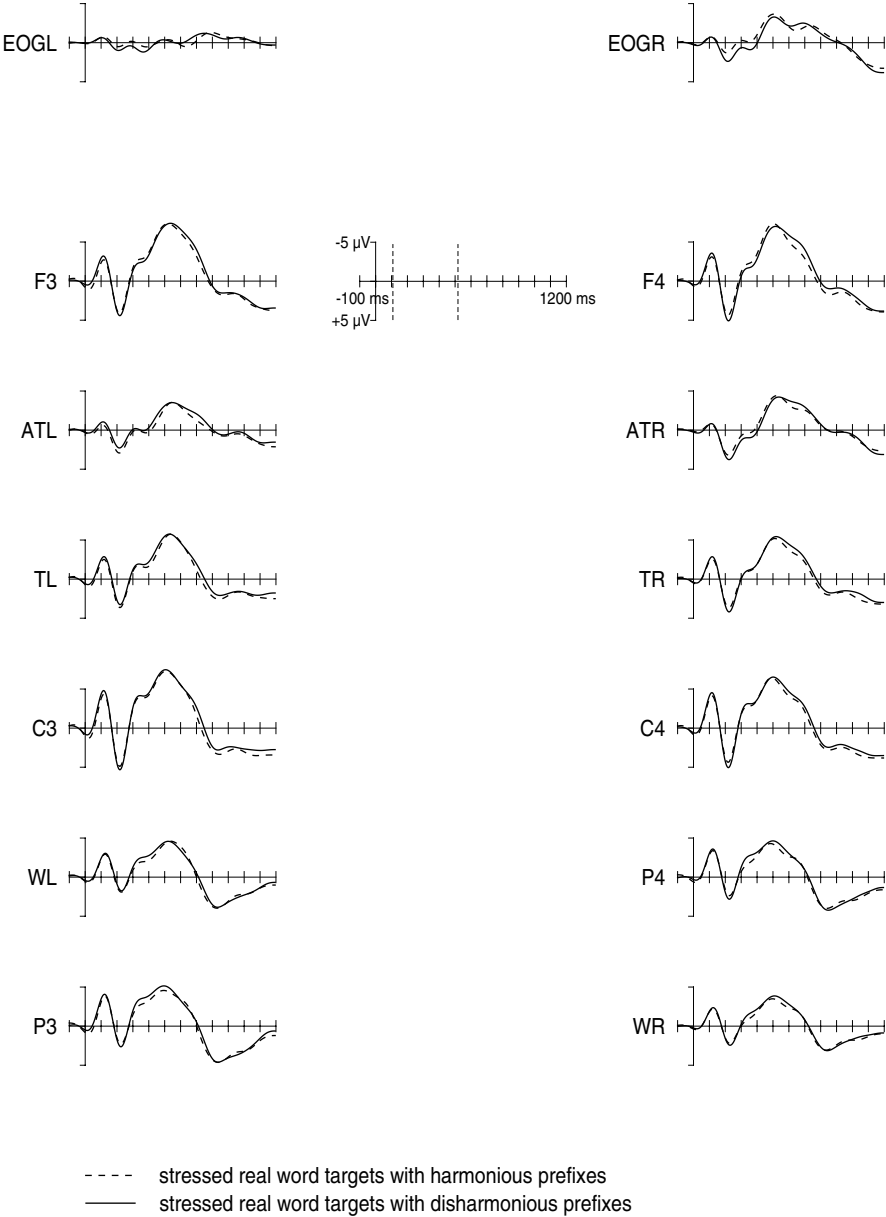


FIGURE 4. ERPs to real word targets preceded by harmonious (broken line) and disharmonious (solid line) prefixes in Experiment 2B (stressed targets). The onset of the nonsense carrier item is at zero milliseconds. Vertical dotted lines mark the acoustic onset and offset of the targets.

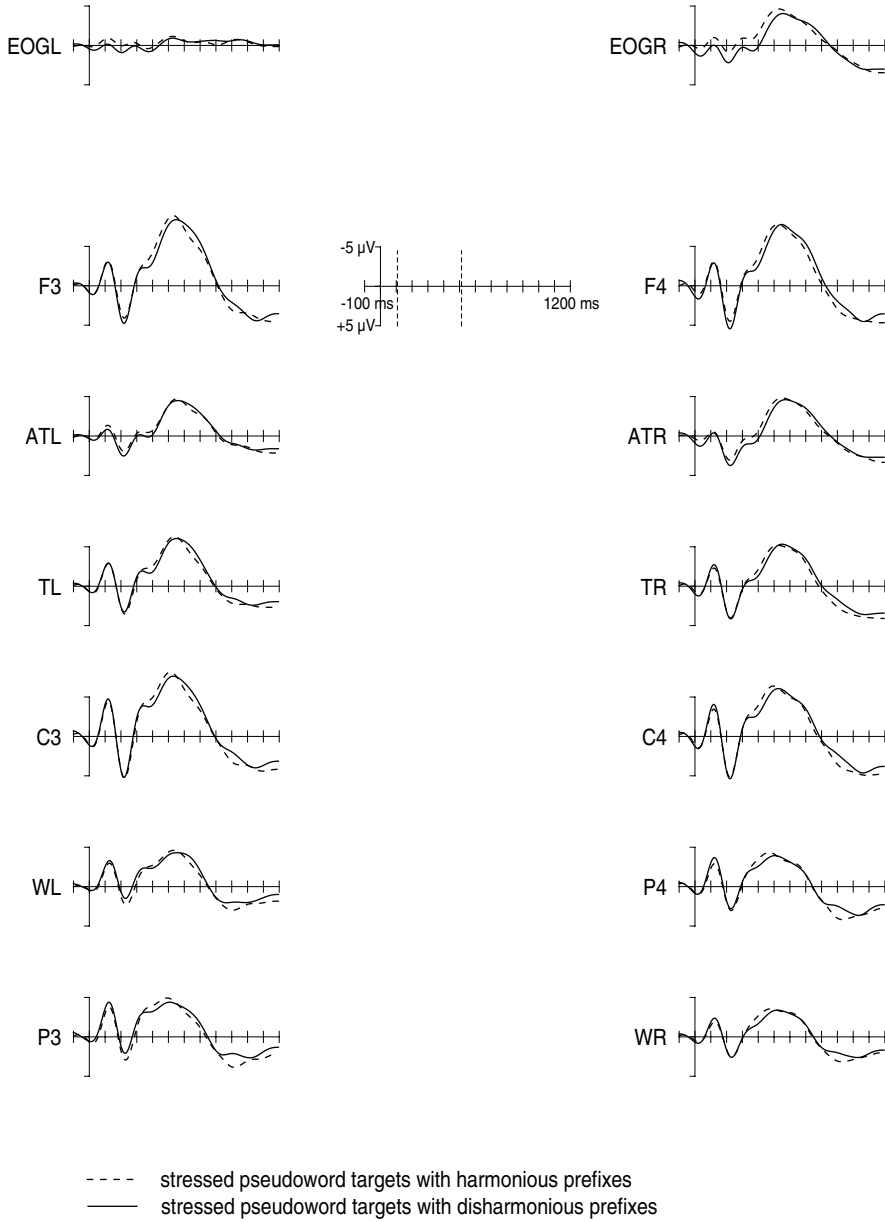


FIGURE 5. ERPs to pseudoword targets preceded by harmonious (broken line) and disharmonious (solid line) prefixes in Experiment 2B (stressed targets). The onset of the nonsense carrier item is at zero milliseconds. Vertical dotted lines mark the acoustic onset and offset of the targets.

In the late time window between 700 and 900 ms Prefix type showed a significant main effect due to more negative ERPs to the disharmonious targets ($F(1,9) = 6.213$, $p = .034$), and also interacted with electrode position ($F(6,54) = 7.180$, $p = .003$, $\epsilon = 389$). The interaction seemed to be caused by more negative response to the disharmonious targets at the centroparietal electrodes. However, no interaction involving Target type and Prefix type was noted.

To summarize, a completely different pattern of brain responses was obtained in Experiment 2B as compared to Experiment 2A. This is due to the change in the stress position. The major differences were related to vowel harmony. In separate analyses in Experiment 2B, no evidence of a vowel harmony effect in the real words was obtained, and in the pseudowords a difference between the targets preceded by harmonious and disharmonious prefixes was noted at around 600 and 700 ms. This is clearly later than the early effects observed in Experiment 2A. In addition, the ERPs recorded to pseudoword targets were more negative than the ERPs to the real word targets especially at the frontal electrodes suggesting that the processing of the stress pattern had a more pronounced effect on the detection of pseudowords. However, we conducted final analyses to find out exactly where the waveforms differed as a function of stress position and vowel harmony. The mean amplitude values in the two ERP experiments measured from the same time windows were subjected to repeated measures of ANOVA similar to the ones in separate experiments, but this time Stress position (first vs. second syllable) was introduced as a between participant factor.

COMPARISON OF ERP TO UNSTRESSED (EXPERIMENT 2A) AND STRESSED (EXPERIMENT 2B) TARGETS

In order to make the visual comparison of the effects of stress and vowel harmony easier, difference waveforms were obtained. The difference waveform showing the stress effect was computed by subtracting (on a sample-by-sample basis) the ERPs to the unstressed targets preceded by the harmonious prefixes (e.g., /PU.katu/) from the ERPs to the stressed items preceded by the harmonious prefixes (e.g., /pu.KAtu/). In this way the processing of the vowel harmony mismatch does not affect the ERPs but instead only the effect of changing the position of the stressed syllable is present. Difference waveforms were created separately for the real and pseudoword targets. (Thus, the more negative the responses to the stressed targets are relative to the unstressed ones, the more negative the resulting difference waveform.) The difference waveforms for selected electrode locations are displayed in Figure 6. For the vowel harmony effect, difference waveforms were created by subtracting the ERPs to the targets preceded by the harmonious prefixes from the ERPs to the targets preceded by

the disharmonious prefixes separately for each experiment and for the real and pseudoword targets. The difference waveforms are displayed in Figure 7. It should be noted that the difference waveforms were low-pass filtered at 6 Hz to remove additional noise due to, for example, alpha activity. Filtering was done for illustrative purposes only and all statistical analyses were performed on the raw data. Finally, as in the previous analyses, normalized amplitude values were used to test differences between conditions in the scalp potential distribution.

Stress position effects. The inspection of Figure 6 shows that the most pronounced effects occur in two time frames, the first one corresponding roughly to the acoustic offset of the target at around 500 ms, and second, occurring in real word targets between 700 and 1000 ms, and in pseudoword targets between 750 and 1200 ms. The ANOVAs performed on the mean amplitudes separately for the different time frames also confirmed these observations. A main effect of Stress position was observed between 700 and 900 ms which was due to more negative responses to the unstressed targets ($F(1,18) = 7.288$, $p = .015$). This suggests that in this time window the processing load of targets with unstressed initial syllables was higher than that of the targets with stressed first syllables, suggesting that the detection of the unstressed targets in Experiment 1 was still going on. The stressed targets were already recognized because this is more or less the time window in which the participants administered button presses to stressed targets in the reaction time experiment 1B (see Table 1 for details).

In the early time window between 400 and 600 ms Stress position interacted with Electrode location and Laterality (both p 's $< .05$). This relates to the complex scalp potential differences in which the processing of the stressed targets involves more activation at the right-sided frontal electrodes as compared to the unstressed targets. The reverse is true at the parietal electrodes; the left hemisphere is more active than the right in the stressed targets.

Overall, stress position had a more prominent effect on the pseudoword targets in the early time yielding a significant interaction of Stress position by Target type between 500 and 700 ms (both p 's $< .05$), and also between 900 and 1100 ms ($F(1,18) = 21.168$, $p < .0001$). In the late window, the difference in the stress effect between the pseudoword and real word targets was most pronounced at the frontal and parietal electrodes as compared to temporal electrodes (Stress position by Target type by Electrode position ($F(6,108) = 5.363$, $p = .006$, $\epsilon = .386$).

Harmony effects as a function of stress. Comparison of the difference waveforms in Figure 7 indicates that two time frames show effects of vowel harmony. Prefix type interacted with Stress position in the late time window between 700 and 900 ms ($F(1,18) = 13.064$, $p = .002$). This was due to more negative ERPs to the harmonious unstressed targets as compared to the disharmonious unstressed targets.

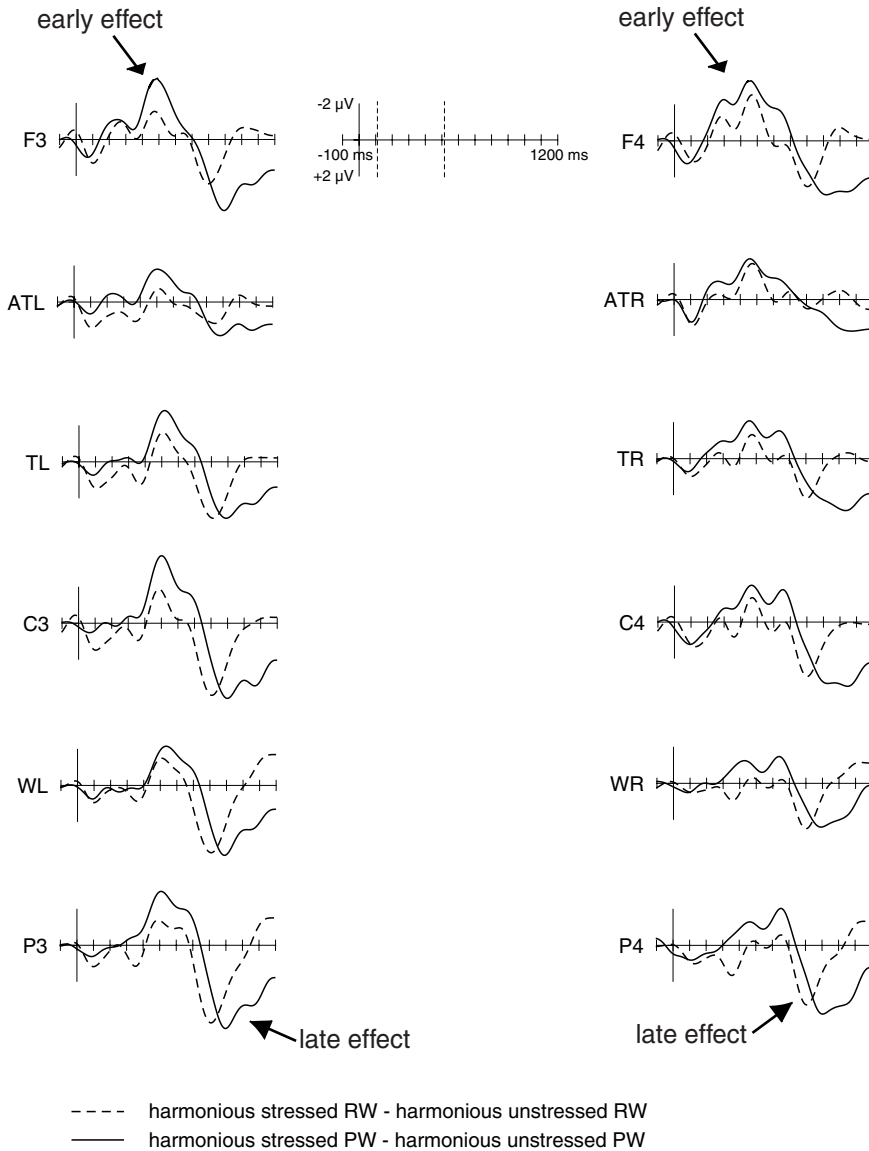


FIGURE 6. Across-experiments comparison; *stress effect*: Difference waveforms obtained by subtracting separately ERPs to unstressed real word (broken line) or pseudoword (solid line) targets preceded by harmonious prefixes from ERPs to stressed targets preceded by harmonious targets. The onset of the nonsense carrier item is at zero milliseconds. Most notable differences are observed in the early time windows between 400 and 600 ms, and in the late time windows between 700 and 1100 ms. The right hemisphere is more active at the frontal electrodes and this effect is more prominent in real word targets. Vertical dotted lines mark the acoustic onset and offset of the targets.

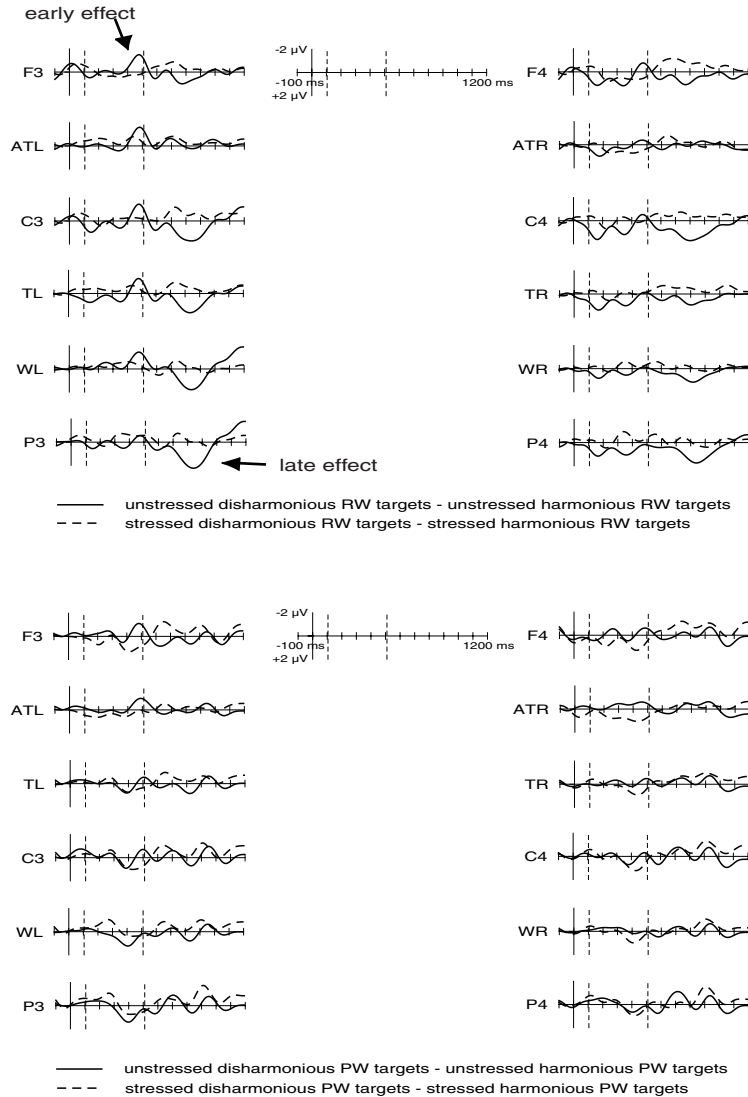


FIGURE 7. Across-experiments comparison; *vowel harmony effect*: Difference waveforms obtained by subtracting ERPs to targets preceded by harmonic prefixes from ERPs to targets preceded by disharmonious targets. This was done separately for Experiment 1 (solid line) and Experiment 2 (broken line), and for real (top panel) and pseudoword (bottom panel) targets. The onset of the nonsense carrier item is at zero milliseconds. Most notable differences are observed only in Experiment 1 (with unstressed targets). The effect is more prominent in real word targets as compared to pseudoword targets in the early time windows between 400 and 600 ms, and in the late time windows between 700 and 1100 ms. Vertical dotted lines mark the acoustic onset and offset of the targets.

Furthermore, this effect was most pronounced at the posterior electrodes resulting in a three-way interaction of Stress position by Prefix type by Electrode position ($F(6,108) = 3.646$, $p = .024$, $\epsilon = .434$). A similar interaction at an earlier time window (between 400 and 500 ms) was also present but this time it was due to more negative responses to the unstressed disharmonious targets as compared to the unstressed harmonious targets at the frontal electrodes ($F(6,108) = 8.197$, $p = .001$, $\epsilon = .386$). A reverse pattern was observed in the stressed targets.

DISCUSSION OF THE ERP RESULTS

The results from the two ERP experiments indicated two separate time windows that showed the most pronounced effects regarding vowel harmony mismatch and stress position. In Experiment 2A, an early time window between 400 and 500 ms presented with more negative responses to the unstressed targets with disharmonious prefixes as compared to the unstressed targets with harmonious prefixes. The later time window, between 700 and 900 ms, showed a reverse pattern. The harmonious unstressed targets yielded now more negative responses than the disharmonious ones. In Experiment 2B, in which the first syllable of the target item was stressed, no reliable differences between conditions were found as a function of prefix type. When the ERPs between experiments were compared, the ERPs to the unstressed targets were more negative than to the stressed targets in the late time window between 700 and 900 ms suggesting that the processing load of the unstressed targets was higher than that of the stressed targets. A simple explanation to account for the results is that both the detection of the vowel harmony mismatch in the unstressed targets and the processing of the stress pattern of the targets with the stressed initial syllable began early, i.e., before the acoustic offset of the targets. However, the prominence level of the stressed syllable could be determined fast, and lexical access (or in the case of pseudowords, an attempt at lexical access) was initiated and the nature of the target was determined rapidly. The larger involvement of the right hemisphere as suggested by a more negative ERP response to the stressed targets at the right hemisphere electrodes is in accordance with results indicating right hemisphere participation in the processing of prosody both in normal participants (e.g., Blumstein & Cooper, 1974; Zatorre, 1988) and in right-hemisphere damaged patients (e.g., Pell & Baum, 1997). In the case of the unstressed targets, the detection of the vowel harmony mismatch seems to be a time consuming process, which was still going on in the late time window. In these time windows both the disharmonious real and pseudoword targets yielded more negative responses than the corresponding stressed targets.

The early and more widely distributed left sided negativity to the unstressed real word targets with disharmonious prefixes as compared to the corresponding pseudoword targets may reflect the fact that information about the phonological code of the real word was made available during the process of vowel disharmony detection. It is possible that the form-based representation is activated and the availability of the phonological code further facilitates detection of vowel harmony mismatch. For pseudoword targets, the early harmony effects were in general weak.

GENERAL DISCUSSION AND CONCLUSIONS

Behavioral and electrophysiological results both revealed significant word stress and vowel harmony effects. Convergence of results was found in that correct stress position on the first syllable of the target stimulus facilitated response speed and yielded earlier ERP effects both in the real and pseudoword targets. Second, vowel harmony mismatch speeded up RTs and showed significant differences in the ERPs, which were more pronounced for the real word targets. Finally, and most importantly, a vowel harmony effect was noted both in the RTs and ERPs only when word stress provided incorrect information about the word boundary. An analysis of the temporal pattern provided by the ERPs further suggested that the processing of vowel harmony mismatch started early, but lasted for several hundreds of milliseconds after the acoustic offset of the stimuli. The most pronounced ERP effects of vowel harmony were related to the late positivity, which is interpreted as a correlate of controlled and strategic processing reflecting subjective evaluation and updating of the context in which the stimuli are perceived (Donchin & Coles, 1988).

The facilitatory effects of word stress can easily be accounted by assuming that the prominence level of a syllable is detected pre-lexically, and as such it provides a cue where to start lexical access. This view is also supported by the current behavioral and electrophysiological data. According to Shortlist (Norris, McQueen, & Cutler, 1995), in English, segmentation at strong syllables will result either in facilitation or slowing down of RTs depending on whether the candidate is aligned with the intended word boundary or embedded in another word. For example, the original demonstration by Cutler and Norris (1988) showed that “mint” is more difficult to recognize when it is embedded in [mɪntɛɪf] as opposed to [mɪntəf]. The explanation is that in [mɪntɛɪf] a word boundary is assumed between the two strong syllables so that both [mɪnt] and [ɛɪf] are segmented. It takes extra time to reconstruct [mɪnt] over the segmentation point. In a similar vein, one can assume that in lexical segmentation of Finnish a stressed syllable will boost up the activation level of each

lexical candidate that is aligned with that syllable, and decrease activation of candidates that are misaligned.

In contrast, however, current results question the assumption that the detection of the vowel harmony mismatch mainly operates on a pre-lexical level (Suomi et al., 1997). The main body of evidence comes from the ERP experiments, which suggest that the processing of the vowel harmony mismatch spans a time interval of several hundreds of milliseconds. Furthermore, the difference between the unstressed and stressed targets was most pronounced several hundreds of milliseconds post stimulus offset, coinciding approximately with the time window in which participants issued a behavioral response to targets in which the first syllable was stressed. This was indexed in the ERPs by the slow positivity, which for the targets with an unstressed first syllable was more negative. In general, this type of positive response is usually correlated with post-lexical or controlled processes as distinct from automatic lexical activation. The surplus negativity to the unstressed targets in this time window indexes higher processing load (Brown et al., 2000), which can be interpreted that the detection of vowel harmony mismatch (or the recognition of the unstressed targets) is still going on.

Based on the ERP data obtained in the current experiments, one cannot unequivocally pinpoint the exact processing level regarding stress and vowel harmony mismatch as cues to word boundaries. This is due to the fact that ERPs recorded from the scalp are a summation of several and sometimes independent sources of electrical activity. Accordingly, the interpretation of the ERP is complicated by the elicitation of overlapping components in the same latency range (e.g., Kutas & van Petten, 1994). Consequently, at a given time point we cannot be sure how many sources contribute to the observed pattern of voltage fluctuation without additional experimental manipulations. Thus, it is possible that the difference in the ERPs noted between the stressed and unstressed items in the late time window between 700 and 900 ms is not (completely) due to controlled or post-lexical processing, but may in part reflect also simultaneous lexical processing most readily indexed in the earlier time window by a negative-going waveform. When the ongoing (negative) activity summates with the positive-going waveform presumably evoked by post-lexical processing it will result in a less positive late wave showing the observed difference. This is a theoretical possibility that cannot be completely discarded on the basis of the current results.

Two further ERP findings were also observed. First, when stress was located on the first syllable of the target items, the potential distribution of the ERPs indicated a right sided dominance as compared to the condition when the first syllable of the target was unstressed. This finding fits well with the literature suggesting that the

right hemisphere is important in the processing of both linguistic and affective prosody (e.g., Blumstein & Cooper, 1974; Ivry & Robertson, 1998; Pell & Baum, 1997; Zatorre, 1988). For example, Ivry and Robertson have proposed a cue-dependent representation of prosodic information in the brain. The theory postulates that low frequencies are mainly processed by the right hemisphere, whereas high frequencies are processed by the left hemisphere. Of the acoustic correlates of prosody, the fundamental frequency (F_0) is contained in the low frequency band of the acoustic signal. Accordingly, processing of F_0 is lateralized to the right hemisphere. F_0 has been shown to be one of the most important acoustic correlates of both word level (lexical) stress and of affective prosody. Another line of evidence regarding the importance of the right hemisphere in the processing of prosody comes from lesion studies. It has been repeatedly shown that right-hemisphere damaged patients are impaired both in the production and perception of both linguistic and affective prosody, although also left-hemisphere-damaged patients are impaired, sometimes even to a greater degree than right-hemisphere patients (Baum, 1998; Schirmer et al., 2001).

The second issue relates to the early effects related to the detection of the vowel harmony mismatch. In the real word targets the ERP response was more pronounced and the scalp potential distribution more wide spread than in the pseudoword targets suggesting that different types of processing were involved. One way to account for this is that for real words the larger recruitment of the (possibly) left hemisphere brain areas may be related to the activation of the phonological code stored in the form based lexical representation. The availability of this representation may assist the detection of the vowel harmony mismatch. In contrast, the ERPs showed only weak and not consistent vowel harmony effects in the pseudoword targets. However, in the RTs a significant vowel harmony effect was also obtained for the pseudoword targets. It is difficult to find a simple explanation why a vowel harmony effect was observed in the RTs but not in the ERPs. This is an area where more research is needed. One should note that this discrepancy between a positive behavioral effect and a negative ERP effect is the only one in the current study. Overall, the behavioral and electrophysiological measures showed a good (positive) correlation.

In summary, the results both from the behavioral data and the brain evoked potentials point to a similar interpretation. In Finnish, listeners prefer word stress over vowel harmony as a cue to word boundary. A simple explanation is that word stress is processed pre-lexically and directly modifies lexical activation. In contrast, the processing of vowel harmony mismatch is a time consuming process, and may involve strategic effects. These results confirm and extend the earlier findings by Vroomen et al. (1998) by providing a more detailed account of the temporal dynamics of the utilization of language specific segmentation cues. These findings

add to the growing body of evidence that the phonological structures of individual languages determine the cues that will be employed in spoken word recognition. Our study is the first to attempt to employ ERPs to investigate the electrophysiological correlates of lexical segmentation of spoken words in continuous speech. Currently, we are far from the full understanding of how spoken word recognition is realized in the brain. However, the present results suggest that the employment of ERPs as a research tool shows a promise in revealing finer details of the temporal characteristics of these processes.

CHAPTER 6

GENERAL DISCUSSION AND CONCLUSIONS

In a series of experiments, the roles of vowel harmony and word stress as cues to word boundaries in spoken Finnish were investigated. The main purpose of the thesis was to find out how multiple cues to word boundaries were employed by listeners, and to reveal in more detail the time course of word recognition in continuous speech. Additional experiments, complementing the main experiments, were run with participants of different language backgrounds in order to explore language specificity of these cues. Another distinctive feature of the current approach is that different methodologies, including reaction time measurement, off-line listening tasks, acoustic analyses and electrophysiological measures were employed. Converging evidence obtained by using different methods will further strengthen the arguments here presented.

The main results of the research reported in this thesis can be summarized as follows:

1. Primary word stress located on the first syllable of the word facilitates word recognition in continuous speech in Finnish (Chapters 2, 4, and 5). This prosodic cue seems to be a language specific feature as listeners with a different language background (such as Dutch and French) did not benefit from word initial stress to the same extent as Finns (Chapter 2).
2. Finnish listeners are sensitive to language specific phonotactic restrictions in detecting words in continuous speech, be it Finnish or an artificial language. In essence, a vowel harmony mismatch occurring at the word boundary facilitates word recognition (Chapters 2, 5). Vowel harmony is a language specific cue as speakers of languages that do not possess vowel harmony restrictions seem to be unable to benefit from this type of a cue (Chapter 2).
3. If multiple cues are present, listeners seem to utilize them so that they prefer one cue over the other. Thus, if word stress provides correct information about the word boundary, listeners will focus on stress and not on vowel harmony. One reason for this seems to be that it takes different amounts of time to process different types of cues. Thus, the earlier a cue is available during the recognition process the more likely it will be utilized. However, if stress information provides conflicting information, or stress is not realized acoustically, listeners will operate with other

available information, such as vowel harmony mismatch. This proposal receives support from electrophysiological data (ERPs), which show that the computation of word stress takes place early on before the acoustic offset of the stimulus word. In contrast, the pattern of data suggests that lexical and post-lexical processes are involved in the detection of vowel harmony mismatch in real words (Chapter 2, and 5).

4. Fundamental frequency (F_0) seems to be an important acoustic correlate of primary word stress in Finnish. This result is based on acoustic measurements of the stimulus materials used in the various experiments. The results might also reflect language specificity in the realization of stress, as for example duration of a phoneme or syllable, which is highly characteristic of the stress pattern in Dutch and English, may not be as important in Finnish. However, this is not to say that differences in duration between stressed and unstressed syllables could not indicate prominence. The acoustic measurements of the stimuli employed by Vroomen et al. showed that duration also co-varied with stress; the stressed syllables were longer than the unstressed ones. As already mentioned, more research is needed to clarify the role of different acoustic parameters as correlates of stress in Finnish (Chapter 3).

5. Listeners seem to be sensitive to fairly subtle changes in the F_0 . Consequently, the acoustic realization of stress in Finnish is not related to sentence level stress, which is usually realized with more distinct acoustic changes. Small pitch movement characteristic to Finnish words in non-accented positions provides enough information to indicate prominence. This may be the basis of a rhythm that helps organize the speech input, which in turn facilitates word recognition (Chapter 4).

HOW DO THE EFFECTS OF WORD STRESS AND VOWEL HARMONY FIT IN A MODEL OF SPOKEN WORD RECOGNITION?

What kind of consequences might these results have on current models of word recognition? Only preliminary and general hypotheses are presented. These could be tested by future research using, for example, simulation studies.

Two alternative accounts of the results are offered. The first one, favored by the current author and one which also gains support from the results of the current thesis, is that, given the temporal differences in the utilization of word stress and vowel harmony, their effects will take place at different stages of the word recognition process. Word stress is probably computed pre-lexically, and thus it directly modifies the activation of lexical candidates. Words that start with a stressed syllable will receive additional activation, which in turn increases their chances to survive in the competition process. The detection of the vowel harmony mismatch,

on the other hand, is a time consuming process, which may be in part under strategic control. This suggestion is based on the slow RTs and also on the ERP pattern in which vowel harmony effects were mainly observed in the late time windows (about 700 to 900 ms postonset of the stimulus).

According to the second alternative, both cues modify lexical processing. However, the degree of modification is related to the relative strength of their role in lexical segmentation. Thus, there is a quantitative, not a qualitative difference between the cues. Word stress is a better cue to word boundary than vowel harmony because it is more widely available. For example, only a vowel harmony mismatch provides information about a word boundary, but word stress informs, in principle, about every word boundary. As already mentioned earlier, results from the ERP study regarding the late time window of the vowel harmony effect is more consistent with the account assuming a post-lexical locus for the vowel harmony effect. The late positivity is in most ERP studies correlated with post-lexical and strategic processing (Donchin & Coles, 1988). Even though indicative, this interpretation is not without doubt. This is due to the fact that ERPs recorded from the scalp are a summation of several and sometimes independent sources of electrical activity. Accordingly, the interpretation of the ERP is complicated by the elicitation of overlapping components in the same latency range (e.g., Kutas & van Petten, 1994). Consequently, at a given time point we cannot be sure how many sources contribute to the observed pattern of voltage fluctuation without additional experimental manipulations. Thus, it is possible that the difference in the ERPs noted between the stressed and unstressed items in the late time window between 700 and 900 ms is not (completely) due to post-lexical processing, but may in part also reflect lexical processing more clearly indexed in the earlier time window by a negative-going waveform. When the ongoing negative activity summates with the positive-going waveform presumably evoked by post-lexical processing it will result in a less positive late wave showing the observed difference. This is a theoretical possibility that cannot be completely discarded on the basis of current results.

Since the word stress effect seems to be pre-lexical, it is reasonable to maintain the idea that word stress directly modifies the lexical activation process. The facilitatory effects of word stress in Finnish could be implemented in a similar manner as the effects of metrical stress in English (McQueen, Norris, & Cutler, 1994) or in Dutch (Vroomen & de Gelder, 1995). A simplistic view could be that the prominence level of a syllable is computed pre-lexically, and any time a stressed syllable is encountered, it can be taken as the first syllable of a word. This is the point where to start lexical access. Thus, the activation level of those lexical candidates that begin with that syllable is boosted up. The extra activation increases the survival rate of the candidates in lexical competition (Norris, McQueen, Cutler, & Butterfield,

1997). The case for vowel harmony is more complicated. Suomi et al. (1997) suggested that the effects of vowel harmony could be modeled in a similar way as the effects of metrical stress; information about the harmony mismatch could increase the activation level of those candidates that are aligned so that the harmony mismatch spans a syllable boundary. Thus, candidates with a higher activation level have a greater chance to survive in the competition process. This is a viable possibility. If so, then vowel harmony effects could be modeled similar to metrical stress in English and word stress in Finnish. However, the main concern with this account is that since the reaction times in general were extremely long in the experiments in which a vowel harmony effect was obtained, it is difficult to assume that such a long time would be spent totally with lexical competition. Furthermore, based on the ERP findings by Tuomainen et al. (Chapter 5) it seems that the detection of a vowel harmony mismatch is a long lasting process starting early during the word recognition process and lasting several hundred milliseconds past the acoustic offset of the target item. The late ERP effects were clearly more pronounced than early effects. This suggests that the detection of a vowel harmony mismatch may not only modulate the competition process, but it will also have an effect on the decision stage. This type of process is not possible to model in Shortlist (Norris, 1994), because it lacks the decision stage, and is only meant to model lexical processing.

If one accepts the view that the long duration of the detection of vowel harmony mismatch relates at least partly to post-lexical processing, then one way to model the facilitatory effects of vowel harmony is to assume a dual route model, in which both the lexical and the phonetic route provide information about the mismatch. The non-lexical route is needed to account for the data obtained with pseudowords. A significant effect was observed in the RTs. In contrast, the ERP results for the pseudoword targets were weak (Chapter 5). It is suggested that for real words lexical representations are also involved in the computation of vowel harmony mismatch. Both of these routes output to the decision system, which will issue a response; in the case of mismatch, a decision about the response can be made earlier. A model that could implement this scenario is Merge (Norris et al., 2000). It contains Shortlist's lexical architecture, which takes care of lexical competition. Furthermore, a decision stage is added to account, for example, for facilitatory effects observed in late occurring targets in pseudowords (Connine, Titone, Deelman, & Blasko, 1997).

SUGGESTIONS FOR FUTURE RESEARCH

Three different lines of future research are suggested all relating to lexical segmentation in one way or the other. The first one concerns the acoustic correlates of

lexical stress. This has to do with the overall applicability of stress in lexical segmentation. In this thesis I have suggested that fundamental frequency is an important acoustic correlate of word stress in Finnish. However, other acoustic parameters are also present but we do not know how they interact as a function of the type of the discourse setting, or whether listeners are sensitive to them.

The second line of research relates to other possible cues to lexical segmentation. A likely candidate based on typological data could be the syllable. Similar to French, Finnish has clear syllable boundaries, and in typological classification is usually considered a syllable-timed language (Karlsson, 1982).

Finally, utilization of the word spotting task as a potential clinical tool needs serious consideration. It might provide a sensitive measure of some of the processes in spoken word recognition. However, construction of a clinical tool is a time consuming task. Before one can employ an experimental method in clinical practice in order to make decisions about individual patients' performance, a representative sample of healthy participants needs to be studied. Another requirement for a task is that it yields a large experimental effect, which is present in a normative sample in (almost) all healthy participants. Based on current findings, this is exactly how word stress and, to a lesser degree, vowel harmony effects behave.

I will elaborate shortly on these three different lines of future research.

Acoustic realization of word stress. As mentioned earlier, research on acoustic realization of word stress in Finnish is minimal. It has been taken for granted, for example, in many textbooks that lexically stressed syllables are also marked acoustically. However, it is clear that not all initial syllables are more prominent than neighboring syllables. So far, acoustic analyses have been performed on different types of read-aloud texts such as TV and radio newscasts. The read-aloud text belongs to a special kind of spoken language, and specific patterns especially regarding the prosody could be inherent in these types of sentences (see for example Mehta & Cutler, 1988). Thus, if one wants to pursue this line of research, corpora consisting of natural spoken language recorded in different settings should be analyzed. The acoustic parameters that should be considered are F_0 , duration, amplitude, and spectral tilt. Finally, the perceptual relevance of the acoustic results should be confirmed with listening tests.

The role of the syllable in Finnish. Syllable boundaries in Finnish are clearly marked. The most frequent syllables have the simple CV or CVC pattern. These features suggest that Finnish could qualify as a language similar to French in which syllable could be *the* unit for segmentation. For some reason, no serious effort has been made to directly test this hypothesis (see, however, Berg & Niemi (2000), and Niemi (in

press), for studies investigating the (internal) structure of the syllable in Finnish). A major difference between Finnish and French is that Finnish has a very rich morphology which yields usually long words with several morphemes attached to the end of the word stem. Consequently, it could be that a syllable based segmentation strategy is not as effective as in French. Because words are long, it could yield an unacceptably high false alarm rate. However, testing of the syllable based lexical segmentation strategy should be easy because an experiment similar to Cutler et al. (1986) is fairly simple to administer.

Clinical applications. Informal observations obtained during testing of a large set of participants suggest that the word spotting task seems to be extremely sensitive to language problems. Preliminary experiences with very mildly aphasic persons and persons with minor reading disorders (suggestive of phonological problems) indicate that they find the word spotting task very difficult to perform. The task by itself could provide a valuable tool in detecting residual language disorders in the processing spoken words. Furthermore, it might be useful in further pinpointing the functional locus of word recognition deficits by using different experimental materials. Normative data on stress and vowel harmony effects are lacking, but inspection of the results of the individual (young) participants suggests that the facilitatory effects of correct stress position as compared to incorrect position are typically so drastic that it should be observable in all healthy subjects. Vowel harmony effects are not obtained in all subjects or the effect is so small that its clinical value may be smaller. Given the shortage of online clinical tools especially in the investigation of the perception and comprehension disorders of spoken language, it seems worthwhile to consider word spotting. Experimental techniques typical of cognitive psychology and psycholinguistics are slowly moving also to clinical settings, but still too often they rely on off-line tasks (e.g. PALPA; Kay, Lesser, Coltheart, 1992; but see Tyler, 1993 for use of on-line task in aphasia research).

To summarize, we have shown convincingly the facilitatory effects of word stress and vowel harmony on lexical segmentation of Finnish. Word stress is preferred for vowel harmony as a cue to word boundary. We have put forth an argument that word stress is computed pre-lexically but vowel harmony effects mainly stem from post-lexical level of processing. This result readily explains the temporal dynamics of these two cues. A modest suggestion on how to model these effects was also presented. The current studies are the first to address the issue of lexical segmentation with diverse methodology using reaction time, off-line listening tasks and electrophysiological measures of brain activity. This approach shows promise, and will be used in the future research on the same topic.

REFERENCES

- BAUM, S.R. (1998). The role of fundamental frequency and duration in the perception of linguistic stress by individuals with brain damage. *Journal of Speech, Language and Hearing Research*, **41**, 31-40.
- BECKMAN, M.E. (1996). The parsing of prosody. *Language and Cognitive Processes*, **11**, 17-67.
- BECKMAN, M.E., & EDWARDS, J. (1994). Articulatory evidence for differentiating stress categories. In: P.A. Keating (Ed.) *Phonological Structure and Phonetic Form: Papers in Laboratory Phonology III*, pp. 7-33. Cambridge: Cambridge University Press.
- BECKMAN, M.E., & PIERREHUMBERT, J.B. (1986). Intonational structure in English and Japanese. *Phonology Yearbook*, **3**, 255-310.
- BERG, TH., & NIEMI, J. (2000). Syllabification in Finnish and German: Onset filling vs. onset maximization. *Journal of Phonetics*, **28**, 187-216.
- BLUMSTEIN, S., & COOPER, W. (1974). Hemispheric processing of intonation contours. *Cortex*, **10**, 146-158.
- BÖCKER, K.B.E., BASTIAANSEN, M.C.M., VROOMEN, J., BRUNIA, C.H.M., & DE GELDER, B. (1999). An ERP correlate of metrical stress in spoken word recognition. *Psychophysiology*, **36**, 706-720.
- BÖCKER, K.B.E., VELDTHUIZEN, I., TUOMAINEN, J., VROOMEN, J., & DE GELDER, B. (submitted). A neurophysiological manifestation of prosodic contrasts in speech perception.
- BROWN, C.M., & HAGOORT, P. (1993). The processing nature of N400: evidence from masked priming. *Journal of Cognitive Neuroscience*, **5**, 34-44.
- BROWN, C.M., HAGOORT, P., & CHWILLA, D.J. (2000). An event-related potential analysis of visual word priming effects. *Brain and Language*, **72**, 158-190.
- COLE, R., & JAKIMIK, J. (1980). A model of speech perception. In R. Cole (Ed.) *Perception and production of fluent speech*, pp. 133-164. Hillsdale, NJ: Erlbaum.
- CONNINE, C.M., TITONE, D., DEELMAN, T., & BLASKO, D. (1997). Similarity mapping in spoken word recognition. *Journal of Memory and Language*, **37**, 463-480.
- CONNOLLY, J.F., & PHILLIPS, N.A. (1994). Event-related potential components reflect phonological and semantic processing of the terminal word of spoken sentences. *Journal of Cognitive Neuroscience*, **6**, 256-266.
- CONTENT, A., KEARNS, R.K., & FRAUENFELDER, U.H. (in press). Boundaries versus onsets in syllabic segmentation. *Journal of Memory and Language*.

- COULSON, S., KING, J.W., KUTAS, M. (1998). Expect the unexpected: Event-related brain response to morpho-syntactic violations. *Language and Cognitive Processes*, **13**, 21-58.
- CUTLER, A. (1976). Phoneme monitoring reaction time as a function of preceding intonation contour. *Perception and Psychophysics*, **20**, 55-60.
- CUTLER, A. (1986). *Forbear* is a homophone: lexical prosody does not constrain lexical access. *Language and Speech*, **29**, 201-220.
- CUTLER, A. & CARTER, D. M. (1987). The predominance of strong initial syllable in the English vocabulary. *Computer Speech and Language*, **2**, 133-142.
- CUTLER, A., & CHEN-H-C. (1999). Lexical tone in Cantonese spoken-word processing. *Perception & Psychophysics*, **59**, 165-179.
- CUTLER, A., DAHAN, D., & DONSELAAR, V. VAN (1997). Prosody in the comprehension of spoken language: A literature review. *Language and Speech*, **40**, 141-201.
- CUTLER, A., MEHLER, J., NORRIS, D., & SEGUI, J. (1986). The syllable's differing role in the segmentation of French and English. *Journal of Memory and Language*, **25**, 385-400.
- CUTLER, A. & NORRIS, D. (1979). Monitoring sentence comprehension. In W.E. Cooper, & E.T.C. Walker (Eds.) *Sentence Processing: Psycholinguistic Studies Presented to Merrill Garrett*, pp. 113-134. Hillsdale, NJ: Erlbaum.
- 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. & OTAKE, T. (1994). Mora or phoneme? Further evidence for language-specific listening. *Journal of Memory and Language*, **33**, 824-844.
- DEACON, D., HEWITT, S., YANG, C-M., NAGATA, M. (2000). Event-related brain potential indices of semantic priming using masked and unmasked words: evidence that the N400 does not reflect a post-lexical process. *Cognitive Brain Research*, **9**, 137-146.
- DELL, F., & VERGNAUD, J. -R. (1984). Les développements récents en phonologie: Quelques idées centrales. In F. Dell, D. Hirst, & J.-R. Vergnaud (Eds.) *Forme sonore du langage*, pp. 1-42. Paris: Hermann.
- DONCHIN, E., & COLES, M. (1988). Is the P300 component a manifestation of context updating? *Behavioral and Brain Sciences*, **11**, 357-374.
- FEAR, B. D., CUTLER, A., & BUTTERFIELD, S. (1995). The strong/weak syllable distinction in English. *Journal of the Acoustical Society of America*, **97**, 1893-1904.
- FRIEDRICH, F.J., HENIK, A., & TZELKOV, J. (1991). Automatic processes in lexical access and spreading activation. *Journal of Experimental Psychology: Human Perception and Performance*, **17**, 792-806.

- FRY, D.B. (1958). Experiments in the perception of stress. *Language and Speech*, **1**, 126-152.
- GÅRDING, E. (1967). *Internal Juncture in Swedish*. Travaux de l'Institute de Linguistique de Lund 6.
- GARNSEY, S.M., TANENHAUS, M.K., & CHAPMAN, R.M. (1989). Evoked potentials and the study of sentence comprehension. *Journal of Psycholinguistic Research*, **18**, 51-60.
- GOW, D.W.,JR., & 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.
- HAGOORT, P., & BROWN, C. (2000). ERP effects of listening to speech: semantic ERP effects. *Neuropsychologia*, **38**, 1518-1530.
- HAHNE, A., & FRIEDERICI, A. (1999). Electrophysiological evidence for two steps in syntactic analysis: Early automatic and late controlled processes. *Journal of Cognitive Neuroscience*, **11**, 194-205.
- HART, J., COHEN, A., & COLLIER, R. (1990). *A Perceptual Study of Intonation*. Cambridge: Cambridge University Press.
- HAYES, J.R., & CLARK, H.H. (1970). Experiments on the segmentation of an artificial speech analogue. In: J.R. Hayes (Ed.) *Cognition and the Development of Language*, pp. 221- 234. New York: John Wiley & Sons.
- HERMES, D. (1997). Timing of pitch movements and accentuation of syllables in Dutch. *Journal of the Acoustical Society of America*, **102**, 4, 2390-2402.
- HERMES, D.J. & RUMP, H.H. (1994). Perception of prominence in speech intonation induced by rising and falling pitch movements. *Journal of the Acoustical Society of America*, **96**, 1, 83-92.
- HIRST, D. & DI CRISTO, A. (1998). A survey of intonation systems. In D. Hirst, & A. Di Cristo (Eds.) *Intonation System*, pp. 1-44. Cambridge: Cambridge University Press.
- HOLCOMB, P.J., & ANDERSON, J.E. (1993). Cross-modal semantic priming: A time-course analysis using event-related brain potentials. *Language and Cognitive Processes*, **8**, 379-411.
- HOLCOMB, P.J., & NEVILLE, H. (1990). Auditory and visual semantic priming in lexical decision: A comparison using event-related potentials. *Language and Cognitive Processes*, **5**, 281-312.
- HYMAN, L.M. (1977). On the nature of linguistic stress. In: L.M. Hyman (Ed.) *Studies in Stress and Accent*, pp. 37-82. Occasional Papers in Linguistics, #4. University of Southern California.
- IIVONEN, A. (1999). Intonation in Finnish. In D. Hirst, & A. Di Cristo (Eds.) *Intonation Systems*, pp. 311-327. Cambridge: Cambridge University Press.

- IIVONEN, A., NEVALAINEN, T., AULANKO, R., & KASKINEN, H. (1987). *Puheen intonaatio*. Helsinki: Gaudeamus.
- IIVONEN, A., NIEMI, T. & PAANANEN, M. (1998). Do F0 peaks coincide with lexical stresses? In: S. Werner (Ed.) *Nordic Prosody: Proceedings of the VIIIth conference*, pp. 141-158. Frankfurt am Main: Peter Lang.
- IVRY, R.B., & ROBERTSON, L.C. (1998). *The Two Sides of Perception*. Cambridge MA: MIT Press.
- JUSCZYK, P.W., HOUSTON, D.M., & NEWSOME, M. (1999). The beginnings of word segmentation in English-learning infants. *Cognitive Psychology*, **39**, 159-207.
- KAGER, R. J. W. (1989). *A metrical theory of stress and destressing in English and Dutch*. Dordrecht: Foris.
- KARLSSON, F. (1983). *Suomen kielen äänne - ja muotorakenne (Finnish phonological and morphological structure)*. Helsinki: WSOY.
- KARLSSON, F. (1987). *Finnish grammar*. Helsinki: WSOY.
- KAY, J., LESSER, R., & COLTHEART, M. (1992). *Psycholinguistic Assessment of Language Processing in Aphasia*. Hove: Psychology Press.
- KLATT, D. (1975). Vowel lengthening is syntactically determined in a connected discourse. *Journal of Phonetics*, **3**, 129-140.
- KUTAS, M., & DALE, A. (1997). Electrical and magnetic readings of mental functions. In: Rugg, M.D. (Ed.) *Cognitive Neuroscience*, pp. 197-242. Psychology Press: Hove East Sussex.
- KUTAS, M., FEDERMEIER, K., & SERENO, M. (1999). Current approaches in mapping language in electromagnetic space. In: C. Brown, & P. Hagoort (Eds.) *The Neurocognition of Language*, pp. Oxford University Press: New York.
- KUTAS, M., & HILLYARD, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, **207**, 203-205.
- KUTAS, M., & VAN PETTEN, C. (1994). Psycholinguistics electrified: Event-related brain potential investigations. In M.A. Gernsbacher (Ed.) *Handbook of Psycholinguistics*, pp. 83-143. San Diego: Academic Press.
- LADEFOGED, P. (1975). *A Course in Phonetics*. New York: Harcourt, Brace, Jovanovich.
- LAINE, M., & VIRTANEN, P. (1999). *WordMill Lexical Search Program*. Center for Cognitive Neuroscience, University of Turku, Finland.
- LEHISTE, I. (1960). An acoustic-phonetic study of internal open juncture. *Phonetica*, **5**, (Supplement 5), 1-54.
- LEHISTE, I. (1972). The timing of utterances and linguistic boundaries. *Journal of the Acoustical Society of America*, **51**, 2018-2024.

- LEHTONEN, J. (1970). *Aspects of quantity in standard Finnish*. Jyväskylä: University of Jyväskylä.
- LIBERMAN, M.Y., & PRINCE, A. (1977). On stress and linguistic rhythm. *Linguistic Inquiry*, **8**, 249-336.
- MARSLEN-WILSON, W. D. (1973). Linguistic structure and speech shadowing at very short latencies. *Nature*, **244**, 522-523.
- MARSLEN-WILSON, W. D. (1984). Function and process in spoken word recognition. In H. Bouma, & D. Bouwhuis (Eds.) *Attention and Performance X*, pp. 125-150. Hillsdale, NJ: Erlbaum.
- MATTYS, S. (1997). The use of time during lexical processing and segmentation: A review. *Psychonomic Bulletin & Review*, **4**, 310-329.
- MATTYS, S., JUSCZYK, P.W., LUCE, P.A., & MORGAN, J.L. (1999). Phonotactic and prosodic effect on word segmentation in infants. *Cognitive Psychology*, **38**, 465-494.
- MCCALLUM, W.C., FARMER, S.F., & POCKOCK, P.V. (1984). The effects of physical and semantic incongruities on auditory event-related potentials. *Electroencephalography and Clinical Neurophysiology*, **59**, 477-488.
- MCCARTHY, G., & WOODS, C.C. (1985). Scalp distributions of event-related potentials: An ambiguity associated with analysis of variance models. *Electroencephalography and clinical Neurophysiology*, **62**, 203-208.
- MCCLELLAND, J. L., & ELMAN, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, **18**, 1-86.
- MCQUEEN, J.M. (1996). Word spotting. *Language and Cognitive Processes*, **11**, 695-699.
- MCQUEEN, J.M. (1998). Segmentation of continuous speech using phonotactics. *Journal of Memory and Language*, **39**, 21-46.
- 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., AND SEGUÍ, J. (1981). The syllable's role in speech segmentation. *Journal of Verbal Learning and Verbal Behavior*, **20**, 298-305.
- MEHTA, G., & CUTLER, A. (1988). Detection of target phonemes in spontaneous and read speech. *Language and Speech*, **31**, 135-56.
- MOULINES, E. & CHARPENTIER, F. (1990). Pitch synchronous waveform processing techniques for text-to-speech synthesis using diphones. *Speech Communication*, **9**, 453-467.
- NAGANO-MADSEN, Y. (1992). *Mora and Prosodic Coordination*. Travaux de l'institut de linguistique de Lund, 27. Lund: Lund University Press.

- NAKATANI, L. H., & DUKES, K. D. (1977). Locus of segmental cues for word juncture. *Journal of the Acoustical Society*, **62**, 714-719.
- NAKATANI, L. H., & SCHAFER, J. A. (1978). Hearing 'words' without words: Prosodic cues for word perception. *Journal of the Acoustic Society of America*, **63**, 234-245.
- NESPOR, M., & VOGEL, I. (1986). *Prosodic Phonology*. Dordrecht: Foris Publications.
- NEVILLE, H.J., MILLS, D.L., & LAWSON, D.S. (1992). Fractionating language: Different neural subsystems with different sensitive periods. *Cerebral Cortex*, **2**, 244-258.
- NIEMI, J. (in press). Onset + rhyme or first mora + end: Intrasyllabic structure in Finnish. In: R.Smyth (Ed.) *Festschrift to Bruce Derwing*.
- 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, & Cognition*, **21**, 1209-1228.
- NORRIS, D., MCQUEEN, J.M., CUTLER, A. (2000). Merging information in speech recognition: Feedback is never necessary. *Behavioral and Brain Sciences*, **23**, 299-370.
- NORRIS, D., MCQUEEN, J.M., CUTLER, A. & BUTTERFIELD, S. (1997). The possible-word constraint in the segmentation of continuous speech. *Cognitive Psychology*, **34**, 191-243.
- NORRIS, D., MCQUEEN, J.M., CUTLER, A., BUTTERFIELD, S., & KEARNS, R. (2000). Language universal constraints on the segmentation of English. *Proceedings of the SWAP Workshop, Nijmegen, The Netherlands*, pp. 43-46. Nijmegen: MPI for Psycholinguistics.
- NYGAARD, L. & PISONI, D. (1995). Speech perception: new directions of research and theory. In J. Miller, & P. Eimas (Eds.) *Speech, Language, and Communication*, pp. 63-96. London: Academic Press.
- OTAKE, T., HATANO, G., CUTLER, A., & MEHLER, J. (1993). Mora or syllable? Speech segmentation in Japanese. *Journal of Memory and Language*, **32**, 258-278.
- 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.
- PELL, M.D., & BAUM, S.R. (1997). The ability to perceive and comprehend intonation in linguistic and affective contexts by brain-damaged adults. *Brain and Language*, **57**, 80-99.
- PIERREHUMBERT, J. (1980). *The Phonology and Phonetics of English Intonation*. Ph.D. Thesis, MIT.
- PRAAMSTRA, P., MEYER, A.S., & LEVELT, W.J.M. (1994). Neurophysiological manifestations of phonological processing: Latency variation of a negative ERP

- component time-locked to phonological mismatch. *Journal of Cognitive Neuroscience*, **6**, 204-219.
- QUENÉ, H. (1993). Segment durations and accent as cues to word segmentation in Dutch. *Journal of the Acoustical Society of America*, **94**, 2027-2035.
- RÖSLER, F., PÜTZ, P., FRIEDERICI, A., & HAHNE, A. (1993). Event-related brain potentials while encountering semantic and syntactic constraint violations. *Journal of Cognitive Neuroscience*, **5**, 345-362.
- RUGG, M. (1984). Event-related potentials and phonological processing of words and nonwords. *Neuropsychologia*, **22**, 435-443.
- SAFFRAN, J.R., ASLIN, R.N., & NEWPORT, E.L. (1996). Statistical learning by 8-month-old infants. *Science*, **274**, 1926-1928.
- SAFFRAN, J.R., JOHNSON, E.K., ASLIN, R.N., & NEWPORT, E.L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition*, **70**, 27-52.
- SAFFRAN, J. R., NEWPORT, E. L., & ASLIN, R. N. (1996). Word segmentation: The role of distributional cues. *Journal of Memory and Language*, **35**, 606-621.
- SAFFRAN, J.R., NEWPORT, E.L., ASLIN, R.N., & TUNICK, R.A. (1997). Incidental language learning: Listening (and learning) out of the corner of your ear. *Psychological Science*, **8**, 101-105.
- SCHIRMER, A., ALTER, K., KOTZ, S., & FRIEDERICI, A.D. (2001). Lateralization of prosody during language production: A lesion study. *Brain and Language*, **76**, 1-17.
- 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.
- SHATTUCK-HUFNAGEL, S., & TURK, A.E. (1996). A prosody tutorial for investigators of auditory sentence processing. *Journal of Psycholinguistic Research*, **25**, 193-247.
- SHIFFRIN, R.M., & SCHNEIDER, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. *Psychological Review*, **84**, 127-190.
- SHILLCOCK, R. (1990). Lexical hypotheses in continuous speech. In G.T.M. Altman (Ed.) *Cognitive models of speech processing: Psycholinguistic and computational perspectives*, pp. 24-49. Cambridge, MA: MIT Press.
- SILVERMAN, K., BECKMAN, M.E., PITRELLI, J., OSTENDORF, M., WIGHTMAN, C., PRICE, P., PIERREHUMBERT, J., & HIRSCHBERG, J. (1992). ToBI: a standard for labeling English prosody. *Proceedings of the Second International Conference on Spoken Language Processing*, Banff, Vol. 2, pp. 867-870.
- SLUIJTER, A.M.C. & VAN HEUVEN, V.J. (1996). Spectral balance as an acoustic correlate of linguistic stress. *Journal of the Acoustical Society of America*, **100**, 2471-2485.

- SLUIJTER, A.M.C., VAN HEUVEN, V.J., & PACILLY, J.J.A. (1997). Spectral balance as a cue in the perception of linguistic stress. *Journal of the Acoustical Society of America*, **101**, 503-513.
- SUOMI, K., MCQUEEN, J., & CUTLER, A. (1997). Vowel harmony and speech segmentation in Finnish. *Journal of Memory and Language*, **36**, 422-444.
- TABOSSI, P., BURANI, C., & SCOTT, D. (1995). Word identification in fluent speech. *Journal of Memory and Language*, **34**, 440-467.
- TRUBETZKOY, N. (1958). *Grundzüge der Phonologie, 2e Auflage*. Göttingen: Vandenhoeck & Ruprecht.
- TUOMAINEN, J., VROOMEN, J. & DE GELDER, B. (1998). The perception of stressed syllables in Finnish. *Proceedings of the Fifth International Conference on Spoken Language Processing (ICSLP'98)*, Sydney, Vol. 5, 2195-2198.
- TUOMAINEN, J., VROOMEN, J., & DE GELDER, B. (submitted). Word stress is an important cue to word boundary in Finnish.
- TUOMAINEN, J., WERNER, S., VROOMEN, J., & DE GELDER, B. (1999). Fundamental frequency is an important cue to word boundaries in spoken Finnish. *Proceedings of the 14th Congress of Phonetic Sciences, San Francisco, August 1-5, 1999*, 921-923.
- TYLER, L.K. (1993). *Spoken Language Comprehension: An Experimental Approach to Disordered and Normal Processing*. Cambridge, MA.: MIT Press.
- UMEDA, N. (1975). Vowel duration in American English. *Journal of the Acoustical Society of America*, **58**, 434-445.
- VÄLIMAA-BLUM, R. (1993). Intonation: A distinctive parameter in grammatical constructions. *Phonetica*, **50**, 124-137.
- VAN HEUVEN, V.J., & VAN DEN BROECKE, M.P. (1979). Auditory discrimination of rise and decay times in tone and noise bursts. *Journal of the Acoustical Society of America*, **66**, 1308-1315.
- VAN PETTEN, C. COULSON, S., RUBIN, S., PLANTE, E., & PEAKS, M. (1999). Time course of word identification and semantic integration in spoken language. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **25**, 394-417.
- VROOMEN, J., & DE GELDER, B. (1995). Metrical segmentation and lexical inhibition in spoken word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, **21**, 98-108.
- VROOMEN, J., & DE GELDER, B. (1997). The activation of embedded words in spoken word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, **23**, 710-720.
- VROOMEN, J., TUOMAINEN, J., & DE GELDER, B. (1998). The roles of word stress and vowel harmony in speech segmentation. *Journal of Memory and Language*, **38**, 133-149.

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. *Journal of the Acoustical Society of America*, **94**(3), Pt. 2. 1880.

ZATORRE, R.J. (1988). Pitch perceptions of complex tones and human temporal-lobe function. *Journal of the Acoustical Society of America*, **84**, 566-572.

APPENDIX 1

Experimental items and prefixes used in Chapter 2 (Experiments 1 and 2)

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

APPENDIX 2

Experimental materials (**high and low frequency items**) used in Chapter 4

High frequency stimuli

Harmonious/disharmonious prefix

tö/to	läpi	'hole'
tö/to	väki	'people'
kä/ka	väri	'color'
tö/to	häätä	'emergency'
ty/tu	mäki	'hill'
pä/pa	kyky	'ability'
pä/pa	jäte	'waste'
tä/ta	käsi	'hand'
pa/pä	katu	'street'
pu/py	koko	'size'
to/tö	kulu	'expense'
po/pö	kuva	'picture'
ta/tä	koti	'home'
pu/py	kone	'machine'
to/tö	kuri	'discipline'
ku/ky	palo	'burning'
ta/tä	pora	'drill'
ko/kö	puhe	'speech'
ku/ky	talo	'house'
pu/py	tapa	'manner'
po/pö	tuki	'support'
ko/kö	tuli	'fire'
po/pö	tulo	'income'
ku/ky	noki	'soot'
ka/kä	halu	'desire'
to/tö	haka	'hook'
pa/pä	salo	'backwoods'
tu/ty	sana	'word'
pu/py	sato	'crop'
tu/ty	satu	'fairytale'
ka/kä	sopu	'harmony'
ku/ky	suku	'family'
pa/pä	sade	'rain'
po/pö	valo	'light'
ka/kä	vapa	'pole'
tu/ty	jako	'division'
ka/kä	juna	'train'
po/pö	lama	'recession'
po/pö	lapa	'shoulder'
ku/ky	lava	'platform'
ku/ky	lasi	'glass'
ta/tä	luku	'number'
ta/tä	lupa	'permit'
ko/kö	raha	'money'

Low frequency stimuli

Harmonious/disharmonious prefix

kö/ko	päre	'splint'
py/pu	käki	'cuckoo'
pö/po	köli	'keel'
pö/po	käpy	'cone'
ky/ku	räme	'pine swamp'
ty/tu	rysä	'fyke'
kä/ka	jyly	'rumble'
tä/ta	säle	'lath'
pu/py	koju	'shack'
to/tö	kulo	'forest fire'
po/pö	kumu	'din'
to/tö	kupu	'crawl'
ta/tä	koje	'device'
ko/kö	kuje	'trick'
to/tö	kumi	'rubber'
to/tö	pahe	'vice'
ka/kä	poni	'pony'
ka/kä	poru	'bawling'
po/pö	puna	'red(ness)'
ko/kö	puru	'sawdust'
ku/ky	tavi	'teal'
ku/ky	tavu	'syllable'
po/pö	tomu	'dust'
to/tö	tuma	'(cell) nucleus'
ta/tä	mono	'skiing shoe'
pa/pä	muki	'mug'
tu/ty	muro	'flake'
tu/ty	muta	'mud'
to/tö	nuha	'sneeze'
ka/kä	havu	'twig'
ku/ky	humu	'whirl'
pu/py	sose	'purée'
ku/ky	suma	'jam'
po/pö	supi	'raccoon dog'
tu/ty	suti	'brush'
ka/kä	vako	'furrow'
pa/pä	valu	'casting'
pa/pä	vana	'wake'
tu/ty	jana	'segment of a line'
po/pö	lato	'barn'
pu/py	loru	'rhyme'
ko/kö	rapa	'sludge'
ka/kä	ropo	'mite'
ta/tä	rahi	'stool'

APPENDIX 3

Experimental materials used in Chapter 5 (Experiments 1A, 1B, 2A and 2B)

Real word target stimuli and the prefixes used in the Experiments. In Experiments 1A and 2A, the first syllable of the carrier received word stress; in Experiments 1B and 2B, the second syllable of the carrier received word stress. The prefix is either harmonious or disharmonious with the target word. The same prefixes were used with the pseudoword targets (see below).

pä/pa	näkö	'sight'			
tö/to	käry	'smoky mell'			
kä/ka	pöly	'dust'	ka/kä	halu	'desire'
pö/po	syvä	'deep'	po/pö	haku	'fetching'
ky/ku	sävy	'tint'	to/tö	haka	'hook'
py/pu	hymy	'smile'	pa/pä	salu	'backwoods'
ky/ku	pyry	'blizzard'	tu/ty	sana	'word'
tö/to	väki	'folk'	pu/py	sato	'crop'
ty/tu	mäki	'hill'	ku/ky	suku	'family'
pä/pa	kyky	'ability'	pa/pä	sade	'rain'
tä/ta	pyhä	'holy'	to/tö	valo	'light'
ty/tu	kynä	'pen'	tu/ty	jako	'division'
py/pu	kylä	'village'	ka/kä	juna	'train'
pä/pa	jäte	'garbage'	po/pö	lapa	'shoulder'
kö/ko	jyvä	'grain'	ku/ky	lava	'platform'
ty/tu	jyvä	'grain'	ta/tä	luku	'number'
pö/po	hätä	'distress'	ta/tä	lupa	'permit'
tä/ta	hätä	'carbon monoxide'	to/tö	kulo	'forest fire'
kä/ka	räme	'pine swamp'	to/tö	kupu	'crawl'
tä/ta	säle	'lath'	ko/kö	kuje	'prank'
py/pu	käki	'cuckoo'	ka/kä	poni	'pony'
pö/po	köli	'keel'	ka/kä	poru	'bawling'
pä/pa	käpy	'cone'	ko/kö	puna	'red'
to/tö	kuja	'alley'	ko/kö	puru	'(saw)dust'
po/pö	muna	'egg'	ku/ky	tavu	'syllable'
ku/ky	koru	'jewel'	po/pö	tomu	'dust'
ta/tä	raju	'violent'	ko/kö	tuma	'nucleus'
pu/py	katu	'street'	tu/ty	muro	'cereal'
pu/py	koko	'size'	tu/ty	muta	'mud'
to/tö	kulu	'expense'	to/tö	nuha	'(common) cold'
po/pö	kuva	'picture'	ka/kä	havu	'(fir) twig'
ta/tä	koti	'home'	ku/ky	humu	'whirl'
ku/ky	palo	'fire'	pu/py	sose	'purée'
ta/tä	pora	'drill'	ku/ky	supi	'raccoon'?
ko/kö	puhe	'speech'	tu/ty	suti	'brush'
ku/ky	talo	'house'	ka/kä	vako	'furrow'
pu/py	tapa	'habit'	pa/pä	valu	'casting'
po/pö	tuki	'support'	pa/pä	vana	'wake'
ko/kö	tuli	'fire'	tu/ty	lato	'barn'
po/pö	tulo	'income'	pu/py	loru	'(nursery) rhyme'
			ko/kö	rapa	'sludge'
			ko/kö	romu	'garbage'

Pseudoword target stimuli and the prefixes used in the Experiments. In Experiments 1A and 2A, the first syllable of the carrier received word stress; in Experiments 1B and 2B, the second syllable of the carrier received word stress

pä/pa	räkö	ka/kä	polu
tö/to	näry	po/pö	muku
kä/ka	syly	to/tö	voka
pö/po	kävä	pa/pä	ralo
ky/ku	pyvy	tu/ty	sunu
py/pu	vämy	pu/py	muto
ky/ku	märy	ku/ky	koku
tö/to	säki	pa/pä	sude
ty/tu	hyki	to/tö	halo
pä/pa	häky	tu/ty	kuko
tä/ta	jähä	ka/kä	pona
ty/tu	jynä	po/pö	kupa
py/pu	kölä	ku/ky	tava
pä/pa	kyte	ta/tä	poku
kö/ko	pöra	ta/tä	hapa
ty/tu	kyvä	to/tö	pulo
pö/po	sätä	to/tö	kopu
tä/ta	pykä	ko/kö	puje
kä/ka	jyme	ka/kä	vani
tä/ta	käle	ka/kä	haru
py/pu	höki	ko/kö	tuna
pö/po	käli	ko/kö	turu
pä/pa	kypy	ku/ky	navu
to/tö	huja	po/pö	lamu
po/pö	sona	ko/kö	kuma
ku/ky	saru	tu/ty	suro
ta/tä	muju	tu/ty	jato
pu/py	vatu	to/tö	laha
pu/py	soko	ka/kä	kuvu
to/tö	palu	ku/ky	tamu
po/pö	tuva	pu/py	kase
ta/tä	hati	ku/ky	sapi
ku/ky	julo	tu/ty	luti
ta/tä	kora	ka/kä	loko
ko/kö	rohe	pa/pä	salu
ku/ky	nulo	pa/pä	sona
pu/py	topa	tu/ty	luto
po/pö	vaki	pu/py	saru
ko/kö	vuli	ko/kö	kupa
po/pö	lalo	ko/kö	hamu

APPENDIX 4.

Reaction times and error rates of additional analyses reported in Chapter 5

Reaction times (RTs) and error rates (in parenthesis) from Experiment 1A and 1B to real word and pseudoword targets. Upper panel displays RTs measured from the onset of the target item (e.g. from /k/ in /pu.katu/), and the lower panel displays RTs measured from the offset of the target item. In Experiment 1A, the stress location was on the first syllable of the target item (e.g. /'PU.katu/, "katu" means 'street'), and in Experiment 1B, the stress location was on the first syllable of the target item (e.g. /'pu.KAtu/). In both experiments, targets items (real word and pseudoword) were preceded by either a harmonious or disharmonious prefix (/pu.katu/ vs. /py.katu/ or /pu.vatu/ vs. /py.vatu/, "vatu" is a pseudoword in Finnish).

EXPERIMENT 1A (RTs measured from the onset of the target item)

<i>Prefix type</i>	<i>Target type</i>	
	Real words	Pseudowords
Harmonious	1157 (26%)	1365 (15%)
Disharmonious	1070 (21%)	1286 (9%)

EXPERIMENT 1B (RTs measured from the onset of the target item)

<i>Prefix type</i>	<i>Target type</i>	
	Real words	Pseudowords
Harmonious	774 (15%)	838 (4%)
Disharmonious	769 (10%)	836 (7%)

EXPERIMENT 1A (RTs measured from the offset of the target item)

<i>Prefix type</i>	<i>Target type</i>	
	Real words	Pseudowords
Harmonious	767 (26%)	974 (15%)
Disharmonious	683 (21%)	900 (9%)

EXPERIMENT 1B (RTs measured from the offset of the target item)

<i>Prefix type</i>	<i>Target type</i>	
	Real words	Pseudowords
Harmonious	370 (12%)	434 (5%)
Disharmonious	357 (10%)	435 (7%)

Samenvatting

Taalspecifieke *cues* voor de segmentatie van gesproken woorden in het Fins: Gedrags- en event-related brain potential-onderzoek

In een reeks experimenten is onderzoek gedaan naar de rol van klinkerharmonie en woordaccentuering als *cues* voor woordgrenzen in gesproken Fins. Het hoofddoel van dit proefschrift was uit te zoeken hoe luisteraars meervoudige *cues* bij woordgrenzen gebruiken, en daarnaast een meer gedetailleerd beeld te geven van het tijdsverloop van woordherkenning in lopende spraak. Ook zijn, bij wijze van complementair onderzoek, aanvullende experimenten uitgevoerd bij deelnemers met een verschillende taalachtergrond om de taalspecificiteit van deze *cues* te onderzoeken. Een ander bijzonder kenmerk van de huidige aanpak is het toepassen van verschillende methodes, inclusief de meting van de reactietijd, off-line luisteropdrachten, akoestische analyses en elektrofysiologische metingen. Het samenbrengen van met verschillende methodes verkregen materiaal zal de aangedragen argumenten verder onderbouwen.

De belangrijkste resultaten van het in dit proefschrift beschreven onderzoek kunnen als volgt worden samengevat:

1. Primaire klemtoon op de eerste lettergreep van het woord vergemakkelijkt woordherkenning in lopende spraak in het Fins (Hoofdstukken 2, 4 en 5). Deze prosodische cue lijkt een taalspecifiek gegeven te zijn, gezien het feit dat luisteraars met een andere taalachtergrond (zoals Nederlands en Frans) minder voordeel hadden van accentuering van het woordbegin dan de Finnen (Hoofdstuk 2).
2. Finse luisteraars zijn, zowel in het Fins als bij een kunstmatige taal, gevoelig voor taalspecifieke fonotactische restricties bij woorddetectie in lopende spraak. In wezen vergemakkelijkt een bij de woordgrens voorkomende *mismatch* in de klinkerharmonie de woordherkenning (Hoofdstukken 2 en 5). Klinkerharmonie is een taalspecifieke *cue*, aangezien sprekers van talen die niet beschikken over restricties in de klinkerharmonie niet in staat lijken te zijn van dit type *cue* te profiteren (Hoofdstuk 2).
3. Op plaatsen waar meervoudige *cues* voorkomen, lijken luisteraars van de *cues* gebruik te maken, waarbij ze de ene cue de voorkeur geven boven de andere. Als de accentuering van een woord juiste informatie geeft over de woordgrens, zal de luisteraar zich derhalve concentreren op de klemtoon en niet op de klinkerharmonie. Een reden daarvoor lijkt te worden gevormd door de omstandigheid dat de verwerking van verschillende typen *cues* verschillende hoeveelheden tijd in beslag neemt. Naarmate een *cue* tijdens het herkenningsproces eerder ter beschikking staat, wordt het waarschijnlijker dat deze wordt gebruikt. Als de klemtooninformatie echter

tegenstrijdige informatie verschaft, of de accentuering niet akoestisch wordt gerealiseerd, zal de luisteraar de overige beschikbare informatie gebruiken, zoals een *mismatch* in de klinkerharmonie. Deze veronderstelling wordt onderbouwd door elektrofysiologische gegevens (ERP's), die aantonen dat de verwerking van woordaccentuering eerder plaatsvindt dan de akoestische offset van het stimuluswoord. In contrast hiermee duidt het gegevenspatroon erop dat lexicale en postlexicale processen betrokken zijn bij de detectie van *mismatches* in de klinkerharmonie bij echte woorden (Hoofdstukken 2 en 5).

4. Fundamentele frequentie (F0) lijkt een belangrijke akoestische correlaat te zijn van de primaire klemtoon in het Fins. Deze conclusie is gebaseerd op akoestische metingen van de stimulusmaterialen die in meerdere experimenten werden gebruikt. Deze resultaten kunnen ook een weerspiegeling zijn van de taalspecificiteit bij de realisatie van accentuering. De duur van een foneem of lettergreep, een belangrijk kenmerk van het accentueringspatroon in het Nederlands en Engels, kan bijvoorbeeld in het Fins minder belangrijk zijn. Dit wil echter niet zeggen dat verschillen in tijdsduur tussen beklemtoonde en onbeklemtoonde lettergrepen niet kunnen duiden op prominentie. De akoestische metingen van de door Vroomen et al. gebruikte stimuli hebben aangetoond dat ook de tijdsduur covarieerde met accentuering; beklemtoonde lettergrepen duurden langer dan onbeklemtoonde lettergrepen. Zoals al eerder is genoemd, is verder onderzoek nodig om de rol van verschillende akoestische parameters als correlaten van de Finse accentuering te verklaren (Hoofdstuk 3).

5. Luisteraars lijken gevoelig te zijn voor vrij subtiele veranderingen in de F0. Als gevolg daarvan is de akoestische realisatie van de accentuering in het Fins niet gerelateerd aan de accentuering op zinsniveau, die normaalgesproken wordt gerealiseerd met meer specifieke akoestische veranderingen. Kleine toonhoogteverschillen, kenmerkend voor Finse woorden op niet-geaccentueerde posities, geven genoeg informatie om prominentie aan te geven. Dit zou de basis kunnen zijn van een ritme dat helpt bij de organisatie van spraakinput, die op haar beurt de woordherkenning ondersteunt (Hoofdstuk 4).