Listeners use vowel harmony and word-final stress to spot nonsense words: A study of Turkish and French

BARIŞ KABAK*, KAZUMI MANIWA*, and NINA KAZANINA**

**University of Konstanz, Germany*
***University of Bristol, United Kingdom*

Abstract

The study explores the role of stress and vowel harmony as cues for speech segmentation. Both in French and in Turkish stress is demarcative, typically falling on word-final syllables. Additionally, Turkish (but not French) has a regular front-back vowel harmony which dictates that all vowels within a word must be either front or back. French and Turkish participants performed a target detection task in which they had to spot nonsense words embedded in a longer auditory string. The results show that word-final stress can successfully signal an upcoming word boundary and is used for speech segmentation by speakers of both languages. In the Turkish group but not in the French group we also found a facilitatory effect of vowel disharmony. We conclude that both vowel harmony and stress can independently signal word boundaries and suggest that listeners can exploit these phonological regularities during speech segmentation.

1. Introduction

The conversion of continuous speech signals into discrete lexical units is one of the fundamental stages of the language recognition process, whereby mental representations of words and other meaningful linguistic units get activated in lexical memory. This task is known to be accomplished effortlessly and automatically by the language user although speech signals notoriously lack unambiguous cues for word or morpheme boundaries. It has been suggested that, on the one hand, word segmentation is to a large extent governed by the principles of lexical recognition. According to this view the language user's mental lexicon constitutes the most conspicuous tool for unpacking the speech signal (e.g., Marslen-Wilson and Welsh 1978; McClelland and Elman 1986; Norris 1994; see Klatt 1989 for a review). On the other hand, sub-lexical cues that are probabilistically associated with particular portions of words as well as sound and rhythmic alternations that typically characterize phonological

JLP 1 (2010), 207–224 1836-6346/10/01–207
DOI 10.1515/LABPHON.2010.010

wordhood in individual languages (Cutler and Norris 1988; Cutler et al. 1986; Johnson and Jusczyk 2001; Saffran et al. 1996a; Saffran 2002; Peña et al. 2002) have also been argued to provide pervasive cues for speech segmentation.

Here we focus on two such sub-lexical phonological cues, edge-most stress and vowel harmony. Regular edge-most stress, which coincides with either the left- or the right-edge of a word, often bears a demarcative (i.e., edge-marking) function, which can provide facilitatory cues for word segmentation (Vroomen et al. 1998). The online use of such phonological cues in adult speech recognition may not be surprising in light of the fact that 7.5 month-old infants are sensitive to the predominant stress pattern in their native language (e.g., Jusczyk et al. 1999; Johnson and Jusczyk 2001; Saffran et al. 1996b) and they can use this information to segment artificial languages (e.g., Thiessen and Saffran 2003). Furthermore, rules that dictate segmental co-occurrence restrictions such as vowel harmony have long been suggested as providing invaluable cues for the learner in the course of language acquisition (Trubetzkoy 1939; Peters 1985) and have been shown to facilitate the detection of word onsets by adults whose native language employs vowel harmony regularities (Suomi et al. 1997). The common denominator of these findings is that language users, infants and adults alike, can effectively apply phonological and/or distributional patterns from their native language when faced with the problem of speech segmentation.

In this article, we investigate the role of stress and vowel harmony as cues for speech segmentation in French and Turkish. In these languages, stress is demarcative, typically falling on word-final syllables. As such, we assume that stress can signal a word boundary immediately after a stressed syllable, and hence it can serve as a facilitatory cue for detecting a target word that comes after such a boundary in these languages. Additionally, Turkish, but not French, has a regular front-back vowel harmony which dictates that all vowels within a word must be either front or back. If vowel harmony is used as a word segmentation cue, we expect its effect to arise in Turkish but not in French. The article is organized as follows. In Section 2, we review previous studies that investigated stress and vowel harmony regularities in word segmentation, and highlight the relevant phonological properties of our test languages, French and Turkish, to lay the foundation for the experimental study. We discuss the design of our target detection experiment and our predictions in Section 3. Section 4 presents the results of the experiment. We address the implications of our findings for word segmentation in Section 5.

2. Stress and vowel harmony regularities in word segmentation

Stress regularities are typically assumed to provide conspicuous cues for word segmentation by the following logic: the culminative nature of accent, which requires that every lexical word has one primary stress, signals the language user the presence of a lexical word for every (primary) stress bearing syllable, indicated by the acoustic/articulatory correlates of stress prominence such as pitch, duration, and intensity (or a combination of these features) among others. This piece of information coupled with the edge-demarcation function of stress in a given language should then signal the language user the existence of word onsets or offsets relative to the position of the stressed syllable in the speech string. More specifically, in a left-most stress language such as Finnish or Hungarian, where stress consistently falls on the initial syllable of a word, the detection of primary stress should automatically signal the beginning of a word. On the contrary, when stress regularly falls on the final syllable of a word, as is the case in right-most stress languages such as Turkish or French, a prominent syllable in the speech should signal the end of a word (or possibly a larger part of speech), thereby allowing the language user to anticipate the onset of a new word. Vowel harmony regulations too can help in demarcating individual words since they dictate which vowel features are (dis) allowed to co-occur within a word. In vowel harmony languages, the span of vowel harmony is often the phonological word, which crucially contains a stem with any associated affixes and arguably coincides with the units that listeners divide speech strings into for lexical access. Exploiting this abstract knowledge, the language user should then be able to postulate a word boundary whenever a sequence of two adjacent vowels is disharmonic.

Suomi et al. (1997) investigated the role of vowel harmony in word segmentation in Finnish where individual (uncompounded) words can either contain the vowels from the set /u, o, a/ (back vowels) or those from the set /y, ø, æ/ (front vowels), with the vowels /i/ and /e/ being free to combine with either set. This sort of co-occurrence restriction is known as palatal (a.k.a. front-back) harmony and is commonly found in Uralic and Altaic languages. In a word spotting task (Suomi et al. 1997, Experiment 1), Finnish listeners heard trisyllabic nonsense words containing a monosyllabic prefix (e.g., *ku*- or *py*-) followed by an existing disyllabic target word with either all back (e.g., *palo* ‘fire’) or all front vowels (e.g., *hymy* ‘smile’). The target word was either harmonious (e.g., *kupalo*; *pyhymy*) or disharmonious (e.g., *pypalo*, *kuhymy*) with the prefix, which, following the stress pattern of Finnish, always bore primary stress. Finnish participants were asked to press a response key immediately upon hearing a real word at the end of a nonsense word. Overall, words in disharmonious strings (e.g., *hymy* in *puhymy*) were detected faster and more accurately than words in harmonious strings although the effect was larger for words with front harmony vowels than those with back harmony vowels. In a similar task (Experiment 2), Suomi et al. asked Finnish listeners to spot real target words, this time at the beginning of trisyllabic strings instead of at the end, containing target words followed by suffixes which were either harmonious (e.g., *hymy-py*) or disharmonious (*hymy-pu*) with the preceding word. Unlike in Experiment 1 of Suomi et al. (1997), no effect of harmony mismatch was found, suggesting a restricted use of harmony information as a segmentation cue detecting word onsets rather than word offsets. Follow-up experiments in Suomi et al. (1997), where the position of target words varied (Experiments 3–4) as well as acoustic differences between words in harmonious vs. disharmonious contexts were controlled for (Experiment 5), replicated the results obtained from Experiment 1. Suomi and colleagues concluded that vowel disharmony facilitates detection of word onsets in online speech processing since harmony mismatch and word boundaries typically coincide in harmony languages such as Finnish. It should be noted that the study had some limitations, as admitted by the authors themselves. Most critically, in about half of the items in the harmonious context condition, the first two syllables could be continued to form real words, hence creating unwanted lexical competition effects, as opposed to disharmonic contexts where no such competition is possible. For example, the onset of a harmonic string such as *kupalo*, *kupa* competes with an existing word *kupari* 'copper', whereas the onset of a disharmonic context like *kypalo*, *kypa*, has no competitor in the Finnish lexicon. Hence, rather than being an effect of harmony, longer response times in harmonic contexts could have resulted from lexical competition.

In addition to vowel harmony, Finnish speakers can also exploit the leftmost (i.e., word-initial) stress regularity in speech segmentation. Vroomen et al. (1998) investigated the relative importance of multiple cues, i.e., the combined effects of vowel harmony and stress, in speech segmentation. In two experiments (Experiments 1 & 2) the authors varied the position of the stress and the presence of the harmony cue which resulted in four variants (Target Word: *hymy*): (i) *pyHYmy*: no harmony clash, stress cue present (on the target word; stressed syllables are capitalized); (ii) *puHYmy*: harmony clash, stress cue present; (iii) *PYhymy*: no harmony clash, stress cue absent; (iv) *PUhymy*: harmony clash, stress cue absent. The authors found that stress was a strong cue for segmentation in Finnish as suggested by faster detection of targets that bore an initial stress (e.g., *puHYmy* or *pyHYmy*) than those that did not (e.g., *PUhymy* or *PYhymy*). The effect of harmony, on the other hand, was only found when the stress cue was absent, thus suggesting that stress was such a strong cue that it attenuated the contribution of vowel harmony. This led Vroomen and colleagues to rank the contribution of word stress above that of vowel harmony in speech segmentation tasks. As in Suomi et al. (1997), this study also made use of real words, which again confounds the role of phonological cues with other lexical considerations. It can be argued, for example, that listeners detect the target faster in conditions were the target bears an initial stress not because they are aware of the stress regularities of their language, but rather because in these conditions the target (*HYmy*), an existing word of Finnish, is acoustically closer to previously encountered exemplars of the same word. Conversely, when an unstressed target word (*hymy*) is less similar to previously encountered examples, the difference could lead to longer response times.

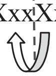
More generally, there is an important consideration that may undermine the claim that Finnish listeners use stress to segment speech. In a stress-initial language the word-initial stress may facilitate lexical access in two ways: (i) it may cue a word-boundary before a stressed syllable, or (ii) it boosts the acoustic saliency of the initial, stressed syllable. Option (i) describes a genuine effect of word-initial stress on speech segmentation, i.e., on dividing a continuous speech input into words. In option (ii), however, the facilitation is not due to a more efficient segmentation, but rather due to an expedited access of the word with a salient onset. A special status of the word onset for lexical retrieval has been one of the keystones of the influential Cohort theory (e.g., Marslen-Wilson 1987). The notion of the importance of word onsets is also adopted by speech segmentation theories such as Cutler and Norris (1988: 118), who argue that "the further information that Word X ends at a point $t + n$ is only of value in that it suggest that lexical unit X + 1 *begins* at $t + n$, but it is irrelevant to the processing of X". Consequently, since primary stress always overlaps with the beginning of the word in Finnish, it is not possible to discriminate between the two options outlined above.

In order to separate the demarcative function of stress from the primacy of word onsets, the current study uses Turkish and French, where stress typically falls on the word-final syllable. Although word stress still has an edge-marking property in these languages, it crucially does not overlap with the onset of a word. As such, while stress can be used to *regressively* postulate a word boundary in a left-most stress language like Finnish, the very same cue might be employed *progressively* in right-most stress languages like Turkish. The difference in the direction of cue-implementation is graphically illustrated in (1) below, where a capital X represents a stressed syllable.

(1) *Left-most word stress (Finnish)*

$Word_1$ ¦ $Word_2$

XxxXxx



Right-most word stress (Turkish, French)

$Word_1$ ¦ $Word_2$

xxXxxX

Hence, we expect word final stress to constitute a useful segmentation cue in both French and Turkish. We also test the role of vowel harmony in speech segmentation. If harmony is used as a word segmentation cue in languages with vowel harmony as part of their phonology, we expect the effect of vowel harmony to arise in Turkish but not in French. Finally, our cross-linguistic study aims to overcome some confounds of previous studies by using nonsense words as targets, which enables us to reach broader generalizations about the use of abstract phonological information in online speech segmentation.

Before we introduce the design of our target detection task in Section 3, we will review the relevant phonological properties of the test languages in question below.

2.1. *Some phonological properties of Turkish and French*

In Turkish, primary stress falls very consistently on the word final syllable. Notwithstanding several exceptions to word-final stress, in the form of loan words, place names, morphologically complex forms with clitic-like suffixes (see Kabak and Vogel 2001, in press, for a review), Turkish stress assignment is highly regular and predictable. As such, Turkish is typically classified as a demarcative stress language, whereby stress regularly signals the right edge of a word. French is also known to employ a regular, quantity insensitive stress assignment algorithm whereby word stress placement can be determined in a straightforward way regardless of the type or length of a word or its internal morphological structure: stress falls on the rightmost full vowel bearing syllable. Consequently, French stress is noncontrastive (cf., among others Dell 1985; Demuth and Johnson 2003; Peperkamp 2004) and carries a demarcative function, systematically denoting the right edge of a word.

As for vowel related properties, Turkish employs an 8-vowel system, with perfectly symmetrical front-back, high-low, and round-unrounded oppositions (i.e., 4 vowels for each feature: front: /i, e, ø, y/, back: /ɯ, a, o, u/; high: /i, ɯ, u, y/, low: /e, ø, a, o/; round: /y, u, ø, o/, unround: /i, ɯ, e, a/). The front-back harmony (a.k.a. palatal harmony) dictates that all vowels within a word should be either exclusively back or exclusively front. As such, within the same word, back vowels are not expected to co-occur with front vowels, and vice versa (e.g., Clements and Sezer 1982; van der Hulst and van de Weijer 1991; Kabak and Vogel 2001, in press). The language also exhibits another type of harmony known as labial or rounding harmony, which is not of concern to the present study. Despite several exceptions to vowel harmony patterns existing in the language, vowel harmony is a highly active process whereby affix vowels obligatorily show alternation depending on the vowel of the host they attach to. As a result, an overwhelming majority of Turkish words, morphologically simplex or simplex alike, are harmonic. Unlike in Turkish, no active phonological process in the form of front-back vowel harmony exists in French, which contains up to 13 vowel phonemes (/i, y, e, ɛ, æ, u, o, ɔ, ø, œ, a, ɑ, ə/).

3. Experiment

3.1. *Participants*

Forty[1] normal-hearing native speakers of Turkish aged 20 to 42 were recruited in Istanbul, Turkey. Forty normal-hearing native speakers of French aged 18 to 35 were recruited from the University of Paris, France, community. All were monolingual speakers and had some experiences in learning nonnative languages (e.g., English, German, Italian and Spanish). Listeners received payment for their participation.

3.2. *Materials*

Thirty-two experimental items consisted of a 5-syllable CVCVCVCVCV auditory sequence with a trisyllabic pre-target string and a disyllabic target (Table 1). The pre-target string and the target were both nonwords in Turkish and French, and were harmonious, i.e., each of them contained only front or only back vowels. However, in half of the cases the pre-target string and the target matched on the frontness/ backness dimension, and thus their concatenation contained only front vowels or only back vowels (harmony-match conditions). In the remaining cases, the pre-target string contained front vowels and the target contained back vowels, or vice versa (harmony-mismatch conditions). Furthermore, the location of stress in the pre-target string was manipulated so that it fell either on the second (penultimate) or on the third (final) syllable of that string (stress-2 vs. stress-3 conditions). Accordingly, there were a total of four conditions arranged in a 2 × 2 design for each of the 32 sets of experimental items: *harmony-match, stress-2*; *harmony-match, stress-3*; *harmony-mismatch, stress-2* and *harmony-mismatch, stress-3*. These 32 sets of experimental materials were distributed across 4 presentation lists using a Latin Square design.

Since we employed a target detection task, each presentation list also contained 224 filler items to ensure an equal proportion of 'Yes/No' responses across all items, an equal number of harmonic/disharmonic targets and an equal probability of a target word occurring in different positions within an auditory string. Ninety-six filler items had harmonic targets, i.e. the vowels

Table 1. *A full sample set of conditions for the target* paVO. *The stressed syllables are in capitals. Syllables with front vowels are in bold.*

	Harmony-match	Harmony-mismatch
Stress-2	goLUshopaVO	**gøLYshø**paVO
Stress-3	goluSHOpaVO	**gølySHØ**paVO

within the target were either both front (e.g., *bøNY*) or both back (e.g., *boNU*). The remaining 128 fillers were disharmonic with the target containing a front and a back vowel (e.g., *baTØ*). Each target was associated with a larger 5-syllable CVCVCVCVCV string which sometimes contained the target and sometimes did not. Whenever the target was contained within the larger 5-syllable CVCVCVCVCV string, it occurred in one of three positions: either as syllables 1–2 (e.g., *koBUsoleTY,* target: *koBU)*, syllable 2–3 (*tøduMEsaFY*, target: *duME*), or syllable 3–4 (*gyNØsuKYtu,* target: *suKY*). Stress was always placed on the second syllable of the target CVCVs, and also on the fifth syllable of the 5-syllable auditory string if the target occupied syllables 1–2 or syllable 2–3 of the auditory string, and on the second syllable when the target occupied syllables 3–4 of the auditory string. Similarly to the experimental items, none of the targets or pre-target strings used for the filler items was existing words of Turkish or French. Special care was paid to ensure that there were no other embedded words in the auditory string and that targets could not be subsets of an existing word of French or Turkish. Thus, there were no effects of lexical competition. There was a total of 96 fillers for which the 5-syllable string contained the target. Of the 96 fillers, 32 contained harmonic targets (all of the 32 experimental items also contained harmonic targets) and 64 disharmonic targets, yielding an equal number of harmonic (n = 64) and disharmonic (n = 64) targets with the expected answer being ‘Yes’. Additionally there were 128 instances where the target word was not part of the corresponding 5-syllable string (hence the expected answer was ‘No’), with an equal amount of harmonic and disharmonic targets.

All stimuli were synthesized based on the voice of a German male speaker (de4: University of Stuttgart) using a diphone-based speech synthesizer (MBROLA; Dutoit et al. 1996). All materials used in the experiment consisted exclusively of consonants and vowels that exist both in Turkish and French (and German). The acoustic properties of our stimuli were based on previously obtained production data from Turkish and French (5-syllable productions; 2 speakers for each group), and we chose values that would be considered natural for these languages while allowing for uniformity in pitch and duration across languages and syllable types regardless of the type of consonant and vowel they contained. More specifically, consonant duration was fixed at 80 ms., and duration for non-stressed and stressed vowels at 80 ms. and 160 ms., respectively, resulting in a total duration of 960 ms. for each CVCVCVCVCV sequence. Fundamental frequency (F_0) for non-stressed vowels was held constant at 116 Hz until 25% of the vowel, linearly interpolated to 110 Hz at 50%, and to 100 Hz at the end of the vowel. Stressed vowels had F_0 values of 124 Hz, 140 Hz, and 75 Hz at these proportions, respectively. An informal auditory inspection of the stimuli by native speakers of both languages confirmed that the values used were reasonable for the intended stress patterns and segmental content.

3.3. *Procedure*

Listeners from each language group were randomly assigned to one of four presentation lists. Participants were seated in a quiet room in front of a computer monitor. They were first presented with a visual target non-word, e.g. *PAVO*, for 500 ms. The visual targets followed the spelling conventions of each language. After a randomized interval of silence ranging from 100 ms. to 500 ms., they heard a 5-syllable nonsense CVCVCVCVCV, e.g. *goluSHOpaVO.* Participants performed a target detection task, i.e., they had to respond as quickly and accurately as possible whether the visually-prompted target occurred within the auditory string by pressing a key labeled as 'Yes' or 'No' on the computer keyboard. For all experimental items the correct response was 'Yes' as the target was always part of the auditory string, whereas the correct response for the filler items was 'Yes' in some cases and 'No' in others. Response times (RTs) for experimental items were originally recorded from the onset of the auditory string and were subsequently adjusted by subtracting the duration of the pre-target CVCVCV string yielding RTs from the target onset.

A short practice session containing 12 items was provided before the experiment to familiarize participants with stimuli and procedure. The experiment was created and controlled using Matlab 7.1 (Mathworks, Inc. 2000) and the accompanying Psychophysics Toolbox extensions (Brainard 1997; Pelli 1997). Auditory materials were presented via Sennheiser HD 250 linear II headphones (for Turkish listeners) or Denon AH-D100 headphones (for French listeners).

3.4. *Predictions*

Our predictions were as follows. If the right-most predictable stress can be used as a cue for predicting a following word-boundary, the target should be detected more easily in the stress-3 conditions, where the stress on the final syllable of the pre-target string cues a word-boundary in between the pre-target and the target, than in the stress-2 conditions, where the stress cues a word-boundary between the second and the third syllable inside the pre-target string. An effect of stress is then expected in both French and Turkish groups yielding faster response times and/or lower error rates in the stress-3 conditions than in the stress-2 conditions. If harmony can be used as a segmentation cue in languages where it is phonologically active, we expect that a harmony-mismatch between the pre-target string and the target will cue a word boundary between them and thus facilitate the target detection for Turkish listeners but not for French listeners. Such a facilitation effect should be reflected in faster response times and/or lower error rates in the harmony-mismatch conditions as compared to the harmony-match conditions in the Turkish group.

4. Results

Statistical analyses were performed on both experimental and filler items, however the latter will not be reported in the current paper for the reasons of space. Statistical analyses of experimental items were performed with accuracy rates and response times as dependent variables. Response times to the experimental items were analyzed by fitting a linear mixed effect model using the *lmer* function from the *lme4* package in R (version 2.6.2; CRAN project; The R Foundation for Statistical Computing 2008). Unlike more traditional ANOVAs, mixed effects models take unaveraged data as input and make it possible to incorporate random effects of both participants and items within a single analysis (for more information on the use of mixed effects models in psycholinguistics see Baayen 2008). Models were fit using a restricted maximum likelihood technique. The model fitting proceeded as follows: initially a model that only included the random factors (participants and items) was applied. This initial model was next enriched by subsequently adding the fixed factors *language* (2 levels: French, Turkish), *harmony* (2 levels: match, mismatch) and *stress* (2 levels: stress2, stress3). Interactions of these factors were then added to the model. Each successive pair of models was evaluated to assess whether the additional factor improved the model fit to the data. The most complex model that significantly improved the fit over the previous model is considered to be the best fitting model and its estimates are reported below. All significant main effects and interactions with $p < .05$ are reported. In all cases where an interaction was significant, we report pairwise comparisons and the relevant 95% confidence interval (95% CI) derived by Markov Chain Monte Carlo simulation (Baayen et al. 2008). A similar procedure was applied for the analyses of the accuracy rates to the experimental items, except a binomial family was used due to a binary nature of the responses (correct/incorrect).

4.1. *Accuracy*

Accuracy rates on filler items were similar between the Turkish and the French groups (Turkish: 85.5%, French: 86.6%). Mean accuracy rates for experimental items for both language groups are presented in Figure 1.

For experimental items, the best-fitting model revealed a significant effect of *language* ($z = -5.7$, $p < .001$), and *stress* ($z = 2.6$, $p = .008$) and *harmony* ($z = 2.3$, $p = .022$). These effects reflected, respectively, that the French listeners were overall more accurate than the Turkish listeners (French: 88%, Turkish: 81%), that targets were detected more accurately in items where the stress fell on the syllable immediately preceding the target than when it fell on the penultimate syllable of the pre-target string (stress3: 86%, stress2: 83%), and

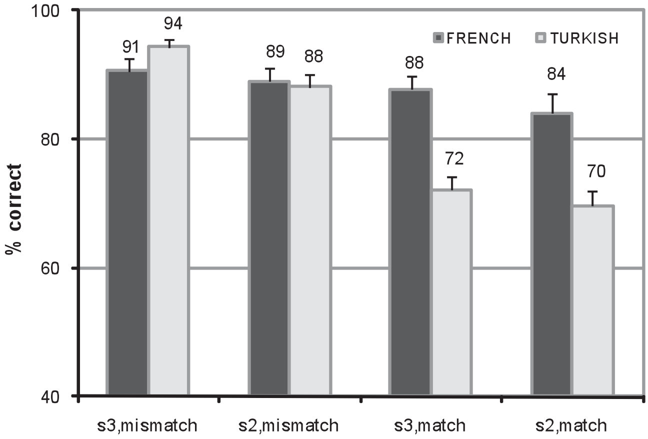


Figure 1. *Mean accuracy rates for French and Turkish speakers in four experimental conditions; s2|s3 = stress falls on the 2nd|3rd syllable of the pre-target string, match|mismatch = the vowels in the pre-target string match|mismatch those in the target non-word on the frontness-backness dimension. 95% error bars from by-participant analysis are shown.*

that target identification was more accurate when there was a harmony mismatch between the pre-target string and the target than when there was not (harmony-mismatch: 91%, harmony match: 78%). Critically, there was also a significant interaction *language* × *harmony* ($z = 5.0$, $p < .001$). Separate analyses on harmony-match and harmony-mismatch items showed that there was a differential effect of *language* in the harmony-match condition (French: 86%, Turkish: 71%; $z = -5.9$, $p < .001$), but not in the harmony-mismatch condition (French: 90%, Turkish: 91%). More pertinently, the significant *language* × *harmony* interaction reflected a clear advantage of harmony-mismatch on accuracy rates in Turkish but not in French (see Figure 1).

Separate analyses within each language group were conducted. ***Turkish:*** In the Turkish group, a significant effect of *harmony* was found ($z = 6.4$, $p < .001$) due to higher accuracy rates in the harmony-mismatch conditions than in the harmony-match conditions (91.3 vs. 70.9% correct respectively). In the same group a significant interaction *stress* × *harmony* ($z = 2.1$, $p = .040$) was also found. Pairwise comparisons within each level of the factor *harmony* in Turkish revealed a significant effect of *stress* in the harmony-mismatch conditions ($z = 2.7$, $p = .007$), but not in the harmony-match conditions ($|z| < 1$). ***French:*** In the French group a significant effect of *harmony* was found (harmony-mismatch: 89.8%, harmony-match: 85.9%; $z = 9.8$, $p < .001$). Note, however, that the size of the harmony effect was much smaller in the French group than in the Turkish group (mean difference between harmony-mismatch and harmony-match conditions; French: 3.9%, Turkish: 20.3%).

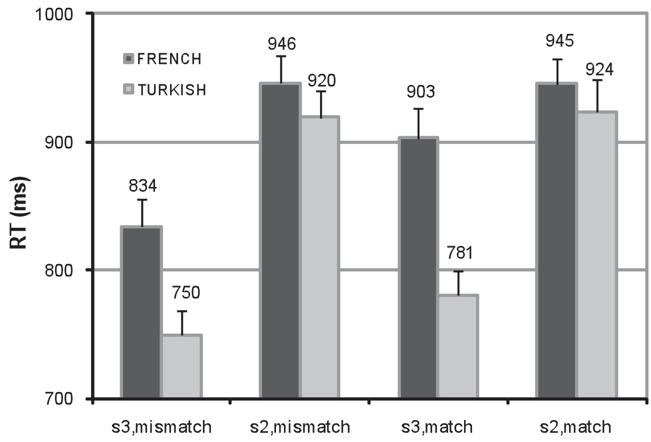


Figure 2. *Mean response times for French and Turkish speakers in four experimental conditions; s2/s3 = stress falls on the 2nd/3rd syllable of the pre-target string, match/mismatch = the vowels in the pre-target string match/mismatch those in the target non-word on the frontness-backness dimension. 95% error bars from by-participant analysis are shown.*

4.2. *Response times (RTs)*

For the RT analyses, incorrectly responded trials were excluded. RTs below 300 ms and those that exceeded a threshold of 2.5 standard deviations above a participant's mean response time for experimental items were replaced by the threshold value. We also excluded the data from 2 French and 2 Turkish participants who responded incorrectly to more than 1/3 of experimental items. Finally, in order to preserve a balanced design, we excluded sets in which individual conditions had an accuracy rate of 10% or lower (1 out of 32 sets were excluded from the French data and 7 out of 32 sets were excluded from the Turkish data). Mean response times in four experimental conditions for both language groups are presented in Figure 2.

The best-fitting linear mixed-effect model applied to the RT data from both languages revealed a significant main effect of all three main factors: *language* ($\beta = -98.3$, $t(1928) = -2.5$, $p = .015$), *stress* ($\beta = -100.6$, $t(1928) = -7.7$, $p < .001$), *harmony* ($\beta = -31.7$, $t(1928) = -2.4$, $p = .016$). [No interactions were significant, i.e., models that additionally included any interactions of these factors showed a poorer fit.]

Planned analyses within each language group were conducted. ***French:*** In the French group, there was a marginally significant effect of *stress* ($\beta = -46.7$, $t(1066) = -1.8$, $p = .067$): targets were identified more quickly in the stress3 condition than in the stress2 condition (stress3: 869 ms, stress2: 946 ms). In additional, a *stress* × *harmony* interaction was found ($\beta = -69.2$, $t(1066) = -1.9$, $p = .051$). Pairwise comparisons within each level of factor *harmony* revealed that the effect of *stress* in French was significant in the harmony-mismatch conditions ($\beta = -116.4$, $t(537) = -4.6$ $p < .001$; mean difference between the stress-2 and stress-3 conditions = 112 ms, 95% CI = ±49 ms), and marginally significant in the harmony-match conditions ($\beta = -47.5$, $t(529) = -1.9$, $p = .055$; mean difference between the stress2 and stress3 conditions = 42 ms, 95% CI = ±49 ms). Cumulatively, the French RT results suggest that stress was used as a segmentation cue by the French listeners. ***Turkish:*** In the Turkish group, a main effect *stress* was found ($\beta = -155.6$, $t(948) = -6.7$, $p < .001$) with shorter RTs in the stress-3 condition than in the stress-2 condition (stress-3: 765 ms, stress-2: 922 ms). [All other main effects and interactions were non-significant in Turkish.]

Hence, the main findings are as follows. On the one hand, harmony had a robust effect on accuracy rates only in the Turkish group, as witnessed by a significant *language* × *harmony* interaction found in the accuracy rates. In Turkish, the target was identified more accurately when there was a harmony mismatch between the pre-target string and the target. On the other hand, the manipulation of the position of stress yielded an effect on response times in both the French and Turkish groups (albeit the effect being marginally significant in French). These results support the claim that stress and vowel harmony regularities that bear demarcative functions can facilitate speech segmentation, and this also applies to non-words.

5. Discussion and conclusion

In this study, we manipulated the position of stress in the pre-target CVCVCV string so that it occurred either on the penultimate syllable (stress-2) or the final syllable (stress-3). Employing two demarcative stress languages, French and Turkish, we hypothesized that stress can signal a word boundary immediately after the stressed syllable, thereby facilitating the detection of a target nonword that follows such a boundary. Consequently, target-word detection was predicted to yield faster RTs and/or lower error rates in the stress-3 condition in both languages. Our results strongly supported this hypothesis. Whenever the target non-word was immediately preceded by a stressed syllable, its detection was faster than when it was not in both languages.

This finding provides critical support for the claim that in languages where stress systematically carries an edge-marking function, stress can be employed as a cue for word boundaries. As noted in Section 2, previous research that addressed this issue (Suomi et al. 1997; Vroomen et al. 1998) examined languages with word-initial stress and could not distinguish whether a facilitatory effect of word-initial stress was due to the postulation of a word-boundary before the stressed syllable or due to the fact that the stress cue overlapped with the word onset thereby making it acoustically more salient. The current study involved languages with word-final stress and thus allowed us to tease apart these possibilities. Our findings demonstrate that stress information in French and Turkish can facilitate speech segmentation by cueing word boundaries despite the fact that it does not promote (arguably, it even demotes) word onsets. An important observation in favor of the claim that predictable edgemost stress can serve as a word-boundary inducer is that it was found to operate both as a regressive, as Suomi et al. (1997) and Vroomen et al. (1998) found, and as progressive cue, as the present study shows. Regressively, it was found to cue a word boundary immediately preceding the stressed syllable in the leftmost-stress language, Finish (albeit caution must be exercised in interpreting this claim, as suggested in Section 2 above). Progressively, the same cue was demonstrated to indicate a word-boundary that immediately follows a stressed syllable in the rightmost-stress languages, French and Turkish.

We also observed that vowel disharmony robustly facilitated target detection in Turkish: targets were detected significantly more accurately in the harmony-mismatch conditions than in the harmony-match conditions. The fact that the effect was robustly present only for the Turkish listeners suggests that language-specific phonological regularities that underlie segmental co-occurrence patterns can influence speech segmentation. In our target detection task both of the phonological cues in question had an effect on Turkish listeners. Previously, Suomi et al. (1997) argued that word stress does not play an important role in speech segmentation in Finnish. In clear contrast to this, Vroomen et al. (1998) argued that stress constitutes a much stronger cue in speech segmentation that ultimately overrides the effect of vowel harmony. They attributed their finding to the fact that an a priori success rate of a vowel disharmony algorithm is inferior to that of a stress-based algorithm. This is because many adjacent words are from the same harmony class in Finnish and as such vowel harmony provides a less reliable and informative cue for word segmentation than stress. Our findings suggest that both cues can have an effect on speech segmentation. We found a significant effect of stress on response times in the harmony-match and the harmony-mismatch conditions for Turkish listeners. Likewise, harmony-mismatch yielded higher accuracy rates in both stress conditions (stress-2 and stress-3). Thus, unlike what was found by Vroomen et al. (1998) for Finnish, we found no signs of attenuation of the effect of one cue by the effect of the other one in Turkish. This is despite the fact that overall in Turkish vowel harmony is a less robust cue to word-boundaries than stress. Similarly to Finnish, the cue is only present in cases where two adjacent words are from different harmony classes and there are no reasons to believe that Turkish syntax employs a systematic rule that would combine words of different harmony types more often than those of the same type. Furthermore, the Turkish lexicon contains a sizeable number of disharmonic words, and some compound words naturally do not obey vowel harmony. As such, it could be argued that, just like in Finnish, Turkish vowel disharmony does not reliably cue word boundaries and that the miss and false-alarm rates for a vowel disharmony algorithm for Turkish might be at least as high as for Finnish reported in Vroomen et al. (1998). Yet, the effect of disharmony was robust in the Turkish group and seems to be independent of the effect of stress, which alone constitutes yet another robust word segmentation cue for this language. However, it is important to note that vowel harmony and stress affected different measures (accuracy rate vs. RTs) thus making it difficult to assess primacy of one cue over the other one.

It should be noted that the fact that French, unlike Turkish, lacks systematic front-back harmony (materialized by vowel co-occurrence constraints in roots or as regular vowel alternations in affixes) does not entail that frontness or backness can never serve as a statistical cue to word boundaries in French. Even in lexicons of non-harmony languages statistical biases based on vowel features might emerge. As noted by an anonymous reviewer, one could imagine a language where some common suffixes that terminate words are more likely to be front than back (or vice versa), thereby increasing the probability of word boundaries following a front vowel. A segmentation algorithm sensitive to such cue is expected to predict, with a certain amount of probability, the presence of a word boundary every time the parser hits a front vowel. [Note, however, the difference between such algorithm and the algorithm based on vowel harmony regularities in harmony languages such as Turkish and Finnish which searches for a *mismatch* of vowel features within two consecutive syllables.] To that end, the marginal effect of vowel harmony found in the accuracy rates of the French listeners may be due to static regularities emerging from the distribution of vowels within French words. An analysis of vowel distribution in French is, however, beyond the scope of this paper.

Finally, our results suggest that although speakers of languages with a fixed stress such as Turkish and French may be 'stress-deaf', i.e., unable to robustly identify the location of the stress in the word or even discriminate between words on the basis of a differential location of stress (Dupoux et al. 1997, 2001; Altmann 2006), they can successfully use the same cue for speech segmentation. A similar situation was found with respect to allophonic and durational regularities that have no phonological significance but can be exploited by speakers in word recognition tasks (Spinelli et al. 2003; Salverda et al. 2003). Arguably, such phonetic cues fail to be substantially and consistently operationalized for identification or discrimination purposes in speech perception tasks (e.g., Werker and Tees 1984; Whalen et al. 1997; Kazanina et al. 2006), suggesting that the status of some phonetic or phonological regularity may be non-uniform across different levels of language representation and processing. Cues that may be insufficient for grouping items in memory (such as predictable stress in French or Turkish) can nevertheless be used in 'instant' processing tasks such as speech segmentation and word recognition to demarcate word boundaries.

To conclude, we showed that listeners can employ, as facilitatory cues to speech segmentation, phonological regularities carrying edge-marking functions. The relative weight of cues emerging from active phonological processes as opposed to those from static regularities remains to be explored.

Acknowledgements

This research was supported by a grant from the German Research Foundation (DFG) within SFB 471 'Variation and Evolution in the Lexicon' to Barış Kabak and a grant from the Royal Society, UK to Nina Kazanina. We would like to thank Prof. Sumru Özsoy who made it possible for us to conduct the study at Boğaziçi University in Istanbul, to Pascale April and Marc Brunelle for their feedback on the French materials, and to Nataliya Chabanyuk, Güliz Güneş and Silke Weber for carrying out the experiments. We are grateful to Heather Goad, John Goldsmith, Paul Iverson, René Kager, Aditi Lahiri, and to the audience of LabPhon-2008 for useful comments and suggestions.

Correspondence e-mail address: baris.kabak@uni-konstanz.de
nina.kazanina@bristol.ac.uk

Note

1. Additional 5 Turkish participants were run, however, their data could not be included due to technical problems that arose during the experiment.

References

Altmann, Heidi. 2006. *The Perception and production of second language stress: a crosslinguistic experimental study*. Newark, DE: University of Delaware dissertation.

Baayen, Harald R. 2008. *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge: Cambridge University Press.

Baayen, Harald R., Doug Davidson & Douglas M. Bates. 2008. Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language* 59. 390–412.

Brainard, David H. 1997. The Psychophysics Toolbox. *Spatial Vision* 10: 433–436.

Clements, George N. & Engin Sezer. 1982. Vowel and consonant disharmony in Turkish. In Harry van der Hulst & Norval Smith (eds.) *The structure of phonological representations* Part 2, 213–255. Dordrecht: Foris Publications.

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

Cutler, Anne, Jacques Mehler, Dennis Norris, & Juan Segui. 1986. The syllable's differing role in the segmentation of French and English. *Journal of Memory and Language* 25. 385–400.

Dell, François. 1985. *Les règles et les sons*. Paris: Hermann.

Demuth, Katherine & Mark Johnson. 2003. Truncation to subminimal words in early French. *Canadian Journal of Linguistics* 48. 211–241.

Dupoux, Emmanuel, Christophe Pallier, Nuria Sebastián-Gallés & Jacques Mehler. 1997. A destressing deafness in French. *Journal of Memory and Language* 36. 399–421.

Dupoux, Emmanuel, Sharon Peperkamp, & Nuria Sebastián-Gallés. 2001. A robust method to study stress 'deafness'. *Journal of the Acoustical Society of America* 110. 1606–1618.

Dutoit, Thierry, Vincent Pagel, Nicolas Pierret, Francois Bataille, Olivier Van der Vreken. 1996. The MBROLA Project: Towards a Set of High-Quality Speech Synthesizers Free of Use for Non-Commercial Purposes. Proceedings of ICSLP'96, Philadelphia, vol. 3, pp. 1393–1396.

Johnson, Elizabeth K. & Peter W. Jusczyk. 2001. Word segmentation by 8-month-olds: When speech cues count more than statistics. *Journal of Memory and Language* 44. 548–567.

Jusczyk, Peter W., Derek M. Houston & Mary Newsome. 1999. The beginnings of word segmentation in English learning infants. *Cognitive Psychology* 39. 159–207.

Kabak, Barış & Irene Vogel. 2001. The Phonological word and stress assignment in Turkish. *Phonology* 18. 315–360.

Kabak, Barış & Irene Vogel. in press. Exceptions to stress and harmony: cophonologies or prespecification? In Horst J. Simon & Heike Wiese (eds.) *Expecting the unexpected: exceptions in grammar.* Berlin: Mouton de Gruyter.

Kazanina, Nina, Colin Phillips, & William J. Idsardi. 2006. The influence of meaning on the perception of speech sound contrasts. *Proceedings of the National Academy of Sciences USA* 103. 11381–11386.

Klatt, Dennis H. 1989. Review of selected models of speech perception. In William Marslen-Wilson (ed.) *Lexical representation and process*, 169–226. Cambridge, Massachusetts: MIT Press.

Marslen-Wilson, William. D. & Alan Welsh. 1978. Processing interactions and lexical access during word-recognition in continuous speech. *Cognitive Psychology* 10. 29–63.

Marslen-Wilson, William. D. 1987. Functional parallelism in spoken word-recognition. *Cognition*, 25. 71–102.

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

Norris, Dennis. 1994. Shortlist: A connectionist model of continuous speech recognition. *Cognition* 52(3). 189–234.

Pelli, Denis G. 1997. The Video Toolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision* 10: 437–442.

Peña, Marcela, Luca L. Bonatti, Marina Nespor, & Jacques Mehler. 2002. Signal-driven computations in speech processing. *Science* 298. 604–607.

Peperkamp, Sharon. 2004. Lexical exceptions in stress systems: Arguments from early language acquisition and adult speech perception. *Language* 80. 98–126.

Peters, Ann M. 1985. Language segmentation: Operating principles for the perception and analysis of language. In Dan I. Slobin (ed.), *The crosslinguistic study of language acquisition-Volume II,* 1029–1068. Hillsdale, New Jersey: L. Erlbaum Associates.

Saffran, Jenny R., Elissa L. Newport & Richard N. Aslin. 1996a. Word segmentation: The role of distributional cues. *Journal of Memory and Language* 35. 606–621.

Saffran, Jenny R., Richard N. Aslin & Elissa L. Newport. 1996b. Statistical learning by 8-month old infants. *Science* 274. 1926–1928.

Saffran, Jenny R. 2002. Constraints on statistical language learning. *Journal of Memory and Language* 47. 172–196.

Salverda, Anne Pier, Delphine Dahan & James M. McQueen. 2003. The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension. *Cognition* 90(1). 51–89.

Suomi, Kari, James M. McQueen & Anne Cutler. 1997. Vowel harmony and speech segmentation in Finnish. *Journal of Memory & Language* 36. 422–444.

Spinelli, Elsa, James M. McQueen & Anne Cutler. 2003. Processing resyllabified words in French. *Journal of Memory and Language* 48. 233–254.
Thiessen, Erik D. & Jenny R. Saffran. 2003. When cues collide: Use of stress and statistical cues to word boundaries by 7- to 9-month-old infants. *Developmental Psychology* 39. 706–716.
Trubetzkoy, Nikolai S. 1939. *Grundzuger der Phonologie*. Travaux du Cercle Linguistique de Prague 7, Prague (English translation in Trubetzkoy 1969).
Trubetzkoy, Nikolai S. 1969. *Principles of phonology*. California: University of California Press, Berkeley and Los Angeles, California. (translated into English from Trubetzkoy 1939).
Van der Hulst, Harry & Jeroen van de Weijer. 1991. Topics in Turkish phonology. In Hendrik Boeschoten & Ludo Verhoeven (eds.) *Turkish linguistics today*, 11–59. Leiden: E.J. Brill.
Vroomen, Jean, Jyrki Tuomainen & Beatrice de Gelder. 1998. The roles of word stress and vowel harmony in speech segmentation. *Journal of Memory and Language* 38. 133–149.
Werker, Janet F. & Richard C. Tees. 1984. Phonemic and phonetic factors in adult cross-language speech perception. *Journal of the Acoustical Society of America* 75(6). 1866–1878.
Whalen, Douglas H., Catherine T. Best & Julia Irwin. 1997. Lexical effects in the perception and production of American English /p/ allophones. *Journal of Phonetics* 25. 501–528.