Pre-print version! Final version to appear in Journal of Phonetics

Acoustic correlates of primary and secondary stress in North American English

Ingo Plag, Gero Kunter & Mareile Schramm

March 24, 2011

Abstract

This study investigates the acoustic correlates of the distinction between primary and secondary stress in English in accented and unaccented morphologically complex words that are either left-prominent or right-prominent (e.g. *ˈvioˌlate* vs. *ˌvioˈlation*). In both accented and unaccented words, the position of primary and secondary stress significantly influences F0, intensity, and spectral balance. The effects are, however, much weaker for unaccented words. A model is presented which can, for accented words, very successfully distinguish the two stress patterns on the basis of pitch, intensity, duration, spectral balance in the two stressed syllables and the pitch slope in the left position. In contrast, the stress patterns of unaccented words cannot be successfully detected on the basis of the acoustic parameters. The findings strongly support an accent-based phonological account of the primary-secondary stress distinction. Primary and secondary stress syllables are not different from each other, unless the word is pitch-accented. In this case what is usually labeled the primary stress syllable becomes the target of a nuclear accent. Left-prominent accented words receive one accent, right-prominent accented words two accents.[1]

1 Introduction

In English and many other languages, the syllables in polysyllabic words are perceived as having different degrees of prominence. These prominence differences within words are referred to as 'stress', and phonologists distinguish between at least three stress-levels: primary, secondary and unstressed (e.g. Hayes 1995). For instance, words such as *ˈnightinˌgale*, *ˌintroˈduction*, or *ˌkangaˈroo* have one syllable that is clearly most prominent and is said to carry primary stress, marked by 'ˈ', one syllable with secondary stress, marked by 'ˌ', and one or more syllables that are unstressed. In phonology textbooks

[1]We would like to thank the anonymous reviewers of this journal and the editor Jennifer Hay for their insightful and constructive comments on an earlier version. Special thanks are due to Stefanie Shattuck-Hufnagel, who generously provided important feedback at an intermediate stage of this project. We are also grateful to Kristina Kösling for collecting some of the data for us, to Jaye Padgett (University of California Santa Cruz) and Parth Bhatt (University of Toronto) for their invaluable help with the data collection. This work was made possible by two grants from the Deutsche Forschungsgemeinschaft (PL151/5-1, PL 151/5-3), which we gratefully acknowledge.

and pertinent reference works (e.g. Giegerich 1992, Hammond 1999, Odden 2005), the phonetic correlates of stress in English are usually given as pitch, intensity, duration and vowel quality, with stressed syllables tending to have higher pitch, higher intensity and longer duration. Furthermore, only stressed syllables can contain the full range of vowel phonemes, while in unstressed position most vowel contrasts are absent.

There is a large body of literature available on the acoustic correlates of stress in English (probably starting with Fry 1955, 1958), and there is a host of parameters that have been suggested to be acoustic correlates of stress. However, previous research has almost entirely focused on the question of how stressed and unstressed syllables differ from each other in terms of their acoustic properties, whereas the distinction between primary and secondary stress has hardly received any attention by phoneticians. This is quite surprising, given that, according to a dictionary-based estimate by Mattys (2000:254), 41 percent of all English words have at least one secondarily stressed syllable.

Furthermore, claims about the acoustic correlates of stress are often based on analyses that do not distinguish between lexical stress on the one hand, and prominence resulting from pitch accent on the other. The notion of pitch accents has taken a central role in intonational phonology, in particular within the framework that has been developed starting with Pierrehumbert (1980). In this framework (see, for instance, Beckman 1994 or Gussenhoven 2004 for thorough treatments), the intonational contour is shaped by the placement of tonal targets, which are assumed to be either low or high for English. Tonal targets can be associated with stressed syllables (pitch accents), or with the boundaries of an intonational phrase (boundary tones and phrase accents). As the presence of a pitch accent lends increased prominence of the respective syllable, different pitch configurations can contribute to the pragmatic meaning of the utterance (see Hirschberg (2004) for details on the interaction between intonation and pragmatics). However, as it will be shown below, the phonetic difference between primarily stressed and secondarily stressed syllables is not crucial to this framework, and it thus remains largely silent about the distinction between these two types of stress.

The series of experiments presented by Fry (1958) are a case in point for the sometimes unclear distinction between accentuation and stress. Fry's experiment investigate the effect that manipulation of F0, duration, and intensity has on the perception of synthetic stimuli such as *subject* or *contract* either as a noun (with primary stress on the first syllable) or as a verb (with primary stress on the second syllable). He finds that the first vowel generally has a lower intensity and a shorter duration if it is not primarily stressed, while changes in the pitch contour are found to be much more important than an absolute F0 difference between the two syllables. Yet, it is not clear whether the differences he finds is merely a difference of lexical stress, or whether the citation form of verbs like *to contract* also differ in their accentuation pattern from that of corresponding nouns. The finding that listeners are particularly sensitive to changes in the pitch contour suggests that the latter is rather likely.

The present study addresses these gaps in the literature by investigating the distinction between primary and secondary stress in accented and unaccented morphologically complex words of English (e.g. *violate, violation, randomize, nominee*), using the acoustic parameters F0, intensity, duration, pitch slope and spectral balance. We present the results of two production experiments that investigate to what extent the five parameters are predictive. The individual effects differ between accented words and unaccented words.

In accented words, left-prominent and right-prominent words show significant differences in F0, intensity, and spectral balance, but not in duration and pitch slope. Unaccented words show the same significant differences, i.e. in F0, intensity and spectral balance, but for F0 and intensity the differences are much smaller. These results partly confirm and partly refute earlier findings from the literature, and underline that effects of stress must be carefully distinguished from effects of pitch accent. We develop models that predict a word's stress pattern (left-prominent or right-prominent) on the basis of the acoustic porperties of the two stressed syllables. The model for accented words is very successful, while for unaccented words the stress pattern cannot be reliably determined on the basis of the acoustics.

The paper is structured as follows. In the next section we provide a review of the literature to prepare the ground for the present study. This is followed by section 3, in which we develop our hypotheses and discuss the methodological details of our investigation. Section 4 presents the analysis of accented words, section 5 the analysis of unaccented words. In section 6 we compare the results of the two experiments, and in section 7 we discuss the significance of our findings and conclude our study.

2 Acoustic correlates of stress in English

Studies such as Fry (1955, 1958), Lieberman (1960), Beckman (1986), Harrington et al. (1998) (see also Laver 1994 for an overview) have shown that there are clear acoustic differences between stressed and unstressed syllables. These differences may concern vowel quality, pitch/F0, duration, intensity, and probably some other parameters, such as spectral balance or high frequency emphasis (e.g. Sluijter and van Heuven 1996a,b). Unfortunately, there are two major problems with many of these results.

First, the vast majority of studies on the acoustic correlates of stress in English have ignored the distinction between primary and secondary stress. For example, Kochanski et al. (2005) analyze all words in their corpus study that have prominence marks (i.e. pitch accents) "on syllables that have primary or secondary lexical stress." (Kochanski et al. 2005:1040) without distinguishing between the two kinds of stressed syllables.

Second, previous studies have suffered from an uncontrolled covariation of accent and stress, or from the fact that they only looked at accented words. In either of these cases, it is unclear whether the effects shown for certain acoustic parameters are effects of accent or effects of stress.

Sluijter and van Heuven (1996a,b) or Oboki (2006) have tried to disentangle the two phenomena and found out that the acoustic differences between stressed and unstressed syllables do depend on accentuation. Sluijter and van Heuven (1996b), for example, find that F0 and intensity cue stress in accented, but not in unaccented words. Similar results presents Oboki (2006), with F0, intensity and amplitude of the first harmonic as pitch accent correlates, and spectral balance, noise at high frequencies and duration as correlates of stress, independent of accentuation.

But even those few studies that included the influence of accentuation suffer from a the first problem mentioned above, the neglect of different levels of stress. As pointed out above, phonologists assume that English has at least three levels of stress, primary, secondary, and no stress. Almost all studies of the phonetic correlates of stress have focused on the differences between stressed and unstressed syllables. Sometimes these studies have also looked at vowel quality, i.e. at the differences between full stressed vowels and reduced unstressed vowels, showing that vowel quality is a reliable cue to stress. Other studies have compared full vowels in main-stressed position with full vowels in other positions, but without asking the question of whether the full vowel in non-primary stress position carries a secondary stress or not.

For example, Oboki (2006) uses stimuli such as *ˈstatue* and *taˈtoo* to investigate the difference between primary stress full vowel syllables and syllables with full vowels that do not have primary stress. As pointed out, for example, by Giegerich (1992), it is controversial whether it is possible at all to determine whether the non-primarily stressed syllable in disyllabic words such as *purist, hotel*, or *July* has a secondary stress, or should be considered unstressed. Some phonological arguments are available for some words, but not for all. E.g. the first syllable of *July* has no stress because it cannot receive primary stress under stress shift (**ˈJuˌly ˈbreakfast*), unlike the first syllable of *ˌhoˈtel* or *ˌchamˈpagne* in contexts such as *ˈhoˌtel ˈmanager* or *ˈchamˌpagne ˈbreakfast*. Whether the non-primary stress full vowels in *statue* and *tatoo* carry secondary stress is unclear, but phonological tests such as stress shift seem to suggest that they are unstressed. Due to this problem, Okobi's results can not be counted as providing clear evidence for the acoustics of primary vs. secondary stress in English.

One of the few studies available of the distinction between primary and secondary lexical stress is concerned with Dutch, Heuven (1987). In that study it was found that intensity, duration and pitch are used to distinguish the two prominence levels. Two studies on English, Fear et al. (1995a) and Mattys (2000a), have investigated the perception of stress differences, but also looked at the associated acoustic correlates, though only in focused, i.e. accented, position.

Fear et al. (1995) look at the full spectrum of primarily stressed vowels vs. secondarily stressed vowels vs. unstressed full vowels vs. unstressed reduced vowels. They find that primarily stressed vowels generally have longer duration than secondarily stressed vowels, but are not statistically different with regard to intensity, pitch and spectral characteristics. Nevertheless, it was found that in an acceptability judgment experiment with cross-spliced stimuli the participants' judgments were predictable on the basis of duration, intensity and spectral measurements, but not pitch. The non-significance of pitch is somewhat surprising, but may be an artifact of the method, i.e. the use of cross-spliced stimuli in an acceptability task, or the small set of words being tested. However, the absence of a significant pitch effect may also be due to the fact that only accented words were used in the Fear et al. study. As Gussenhoven (2004: 20) argues, pitch differences are typically associated with the difference between accented and unaccented syllables, and Fear et al.'s findings would support this argument, as all words they look at are accented. Gussenhoven (2004) argues that there may also be a durational difference between these two stress levels, but he concedes that solid empirical evidence for this is still lacking.

Mattys (2000) used pairs of words that differed in their stress patterns but shared the same root (e.g. *prosecutor* – *prosecution*). The participants in Mattys' experiment were presented with word fragments (the first or the first two syllables of a word) and had to guess whether the fragment belonged to the word with initial primary stress (e.g. *prosecutor*) or the word with initial secondary stress (e.g. *prosecution*). The acoustic analysis of pitch, intensity and duration showed that in left-prominent words (such as *prosecutor*) the first syllable had a much higher pitch than the second (unstressed) syllable, while in right-prominent words (such as *prosecution*) the first syllable did not contrast in pitch with the second (unstressed) syllable. With regard to intensity, no differences between the first two syllables were found, neither with left-prominent nor with right-prominent words. A comparison of primary stress and secondary stress in the first syllable (across words) yielded significantly higher values of pitch and duration, but not of intensity for the primarily stressed syllables. However, in an analysis of the participants' responses in the task, it turned out that the responses could be predicted on the basis of pitch, duration and intensity. Note, however, that the acoustic correlates of primary and secondary stress within a given word were not investigated directly, and it is unclear how the results of Mattys' study relate to this distinction. Furthermore, we learn nothing about the second stress in the word (counted from the left).

Current metrical phonology proposes a different account of the different levels of prominence in English (cf. de Jong et al. 1993, Beckman and Edwards 1994, de Jong 2004, Gussenhoven 2004). Here, stress is seen as a binary feature: a given syllable can be either unstressed or stressed. In addition to that, it is assumed that pitch accents can only be associated with stressed syllables, by which an even higher degree of prosodic prominence is introduced. Thus, this system can account for a three-way distinction of stress levels in a given syllable: unstressed, stressed (but unaccented), and accented (which entails stressed). Thus, what has been usually assumed to be "secondary" and "primary" stress is relegated to the presence or absence of an accent on a stressed syllable. For instance, de Jong (2004:494) considers a stressed, unaccented syllable as bearing "secondary", and a stressed, accented syllable as bearing "primary" stress.

The account given by Gussenhoven (2004:21) is similar in that he defines primary stress as the last stressed syllable in an intonational unit that receives a pitch accent. Yet, he argues that there may be a phonetic difference between syllables with primary and secondary stress even if neither of them is accented. There seems to be strong empirical support for a durational difference between these degrees of stress in Dutch, The findings reported in de Jong (2004) may be seen as experimental support that this difference is also found in English. Here, the duration of pairs such as *bet* and *sporting bet* were compared in an accented and an unaccented environment. The vowel in *bet* was found to be notably longer than that in *sporting bet*, even if the word occurred in post-nuclear position in which no accent was to be expected. However, all stimuli with secondary stress in de Jong (2004) featured the syllable as the second element of a compound.

There are indeed some studies available of the prominence relations in compounds that may be instructive for work on the primary-secondary stress distinction in general. Compounds may be left-prominent (as for example *ˈcase ˌmanager*), or right-prominent, as in *ˌhome ˈphone*, and compound prominence also involves the prominence relation between two stressed syllables, and not between a stressed and an unstressed syllable. Studies of the phonetic correlates of compound stress (e.g. Farnetani et al. 1988, Plag 2006, Kunter and Plag 2007, Kunter 2010) have shown that F0, intensity, duration, spectral balance and pitch slope are indicators of compound prominence, with F0 probably being the most important one. It is unclear, however, whether these results carry over to non-compound words, especially since the compounds tested in all of these studies can be assumed to carry at least one pitch accent.

In how far findings on compounds relate to the acoustics of primary and secondary stress in non-compound words is unclear, as is the question which acoustic correlates contribute to this phonological difference. These gaps are filled by the present study.

To summarize, the acoustic correlates that distinguish primary and secondary stress in English in acccented and unaccented position are essentially still unknown. The present study is the first investigation for English that explicitly and systematically addresses the question of how the five acoustic parameters mentioned above vary between primary and secondary lexical stress within a larger set of naturally spoken data. Furthermore, it differentiates between the position of the respective stresses within the word, and it looks at the effect that accentuation has on the acoustic correlates of stress.

3 Methodology

3.1 Experimental design and participants

We conducted two experiments, each with the same set of target words. Experiment 1 was designed to elicit accented pronunciations; experiment 2 was designed to elicit unaccented pronunciations. In both experiments, the potentially intervening effect of boundary tones or phrase accents was avoided by embedding the target word into carrier sentences. In particular, the target word was followed by "again", which was expected to carry all boundary-related tonal elements.

Accented pronunciations were elicited by having participants read out a carrier sentence with the pertinent item in focus position ("She said 'X' again", 'X' standing for the target word). In experiment 2, participants had to read the same target words, but embedded in different carrier sentences. For each target word, participants had to read out a small dialogue in which the item of interest occurred in the second turn, in the middle between two contrastively focused items, with a distance of two words on either side. A two-word distance of the target word from the pitch-accented word in the phrase has also been used, for example, by Okobi (2006) in his experiments. In such an environment the phrase-level accent is far enough away from the target word so that no effects of the accent will be observed there. An example of such a mini-dialogue is given in (1). Only the second realization of each test item was analysed.

(1) Did PETER say 'X' again?
No, it was JOHN who said 'X' again, not PETER.

Target words were three- to five-syllable complex words that, uncontroversially, have two prominent syllables (unlike some disyllabic words as discussed in the previous section). We will refer to the first prominent syllable occurring in the word as "left position" and to the prominent syllable that occurs later in the word as "right position". Target words were considered either as left-prominent (if primary stress was in left position and secondary stress in right position), or as right-prominent (if secondary stress was in left position and primary stress in right position).

Left-prominent words involved the suffixes -ˌ*ate* or -ˌ*ize* (as in ˈ*random*ˌ*ize*, ˈ*activ*ˌ*ate*). Right-prominent words ended in the suffixes -ˈ*ation* or -ˈ*ee* (as in ˌ*vio*ˈ*lation*, ˌ*publi*ˈ*shee*).

We tested 66 word types with eight different left vowels and three different right vowels. Of the 33 word pairs, 21 contained pairs of the same base with the suffixes -ˌ*ate* and -ˌ*ation*, e.g. ˈ*vio*ˌ*late* vs ˌ*vio*ˈ*lation*. These 21 pairs thus had segmentally identical stressed syllables, but different prominence patterns. A full list of the 66 target words is given in the appendix.

Each experiment involved a practice session to ensure that participants had understood the instruction to "read out the following sentences carefully and as naturally as possible". Items were presented in pseudo-randomized order. Pseudo-randomization involved redistributing items in those cases where, after randomization, more than two items with the same suffix followed each other.

Overall, 40 native speakers of North American English participated in the two experiments. Experiment 1 was carried out with 8 male and 11 female speakers, all students at the University of California, Santa Cruz. The data for experiment 2 come from 9 male and 12 female speakers, all students at the University of Toronto. All recordings were made in a sound-proof booth. Overall, we obtained 2640 word tokens for the acoustic analysis.

3.2 Acoustic measurements

All acoustic measurements were taken using the speech analysis software Praat (Boersma and Weenink, 2010). In each word, the syllables with primary and secondary stress were manually identified, and the sonorant part of the rime of each syllable was annotated as the measurement interval. Then, a Praat script was used to obtain all acoustic measurements used in this study: F0, pitch slope, intensity, duration, and spectral balance.

Following Campbell and Beckman (1997) and Štekauer et al. (2007), F0 was measured as the average F0 throughout the measurement interval. The script used the Praat autocorrelation algorithm, with F0 settings adjusted to suitable settings for male (75–300 Hz) and female (100–500 Hz) speakers. If the tracker generated suspicious F0 contours using these settings (e.g. octave jumps, large stretches without F0 values), the upper and lower boundary were lowered automatically up to three times. If no plausible F0 could be obtained after three automatic adjustments, the observation was discarded. The measurements were subsequently transformed to semitones relative to the lowest observed frequency using the logarithmic transformation $f_{i,\mathrm{ST}} = 12 \cdot \log(\frac{f_i}{\min f}) \cdot \log(2)^{-1}$, where f_i is the i-th average F0 measurement, and $\min f$ the minimum of all observed average F0 measurements (all in Hz).

While average F0 seems to be an appropriate way to assess the overall influence of F0 on different levels of stress and prominence, it is still a simplification of the F0 contour within a given vowel: in principle, a falling, a level and a rising contour may all have the same mean F0. However, Hermes and Rump (1994) and Terken and Hermes (2000) have shown that there are differences in perceived prominence of falling and rising pitch contours. It is thus conceivable that English makes use of these perceptual differences to realize different degrees of stress. Therefore, pitch slopes, i.e. the slope of a line drawn between the F0 maximum and the F0 minimum, were also calculated as a way to capture the difference between falling and rising pitch contours. Thus, for any measurement interval, pitch slope (in ST/s) is derived as $S = \frac{f_{max} - f_{min}}{t_{max} - t_{min}}$, where f_{max} and f_{min} are the F0 maximum and minimum in the interval, and t_{max} and t_{min} are the times at which the maximum and minimum pitches are observed.

Duration was directly obtained from the length of the measurement interval of each syllable under investigation. Intensity (in dB SPL) was retrieved from Praat intensity objects, using Kaiser analysis windows with shape parameter $a = 20$ and the same pitch floor as used in the pitch measurement. The measurements for spectral balance were obtained in a procedure similar to that applied by Sluijter and van Heuven (1996b). A long-term average spectrum of each measurement interval with a bandwidth of 100 Hz was split into two disjoint frequency bands, and mean intensities were taken for each band. The present measure used a high frequency band ranging from 1,000 to 4,000 Hz and a low frequency band ranging from 0 to 1,000 Hz. Spectral balance (in dB SPL) was then calculated as $B = I_{high} - I_{low}$. In general, B should be negative for all measurements, as the source-filter model predicts a lower amplitude of higher-order harmonics (cf. Johnson 2003:80). A negative number closer to zero indicates a largely balanced distribution of energy across the spectrum, while a negative number farther away from zero indicates that the amount of high-frequency energy is particularly low.

3.3 Statistical analysis

For our models in the analysis, we fitted several mixed-effects regression models using the statistical package R (R Development Core Team 2007) and the lme4 package Bates et al. (2007). Depending on what we investigated, the models contained a selection of the following variables of interest: F0, INTENSITY, DURATION, PITCH SLOPE, SPECTRAL BALANCE, POSITION (with the values `left` and `right`) and STRESS (with the values `left-prom` for left-prominent words and `right-prom` for right-prominent words). Where appropriate, these variables were supplemented by suitable control variables such as GENDER, the vowel in left position VLEFT, and the vowel in second position VRIGHT. The models we present have been obtained using the standard simplification procedures, according to which non-significant predictors are eliminated in a step-wise evaluation process (e.g. Baayen 2008). Furthermore, unless otherwise indicated, data points were removed that showed residuals larger than 2.5 standard deviations in the initial fit.

A further note is in order with regard to our inclusion of gender. The analyses we present include gender as a main effect with no further interactions to control for the differences in F0 between male and female speakers. We also devised models with gender interacting with other predictors (e.g. stress and position) in the models. Sometimes we found interactions for gender, but in all these cases the effect of position and stress was going in the same direction for both genders, with the effect only being significantly stronger for one of the genders. Given that in these cases the direction of the effect was the same for both genders, and given that we are not primarily interested in gender differences, we use gender only as a main effect in the main body of the paper.

Following Baayen et al. (2008), speaker and item identifications were included as random intercepts (SPEAKER and ITEM, respectively). For each mixed-effects model, the appropriateness of the random effects structure was verified using likelihood ratio tests. Only those random effects were retained that yielded a significant increase in log likelihood (cf. Baayen 2008:253).

The inclusion of speaker as a random effect controls for individual differences between speakers with regard to speech rate, height of voice, and other, mostly unknown, sources of speaker-specific variability. The random effect ITEM is used to account partially for variation introduced by vowel-intrinsic differences. It is a well-known observation that different vowel phonemes have different intrinsic acoustic characteristics (see, for example [references to Fairbanks and House 1950, House and Fairbanks 1953, Whalen and Levitt 1995]), which may impair the predictive power of a statistical model that estimates prominence levels on the basis of these acoustics.

We verified that the inclusion of ITEM as a random effect can be used to reduce this detrimental influence in the following way. We included the vowels in the left and right positions (VLEFT and VRIGHT) as two additional covariates in the initial models, and fitted these models both with and without the random effect ITEM. When this random effect was included, VLEFT and VRIGHT turned out to be non-significant for F0 and intensity, while they were significant if ITEM was not used as a random effect. These results show that, at least to some extent, including item as random effect can alleviate the problem of vowel-intrinsic differences in these two parameters.

For duration however, things look different – vowel duration was significant even under presence of ITEM as a random effect. Apparently, differences in intrinsic vowel length are so large that they cannot be adequately be expressed by means of a random effect, which underlies certain mathematical restrictions (see Baayen et al. 2008 for details). The methodological implications and solutions to the problem of variable vowel length are discussed in more detail in section 4.

For the models in sections 4.1 and 5.1, we provide tables that document the estimated coefficient for each predictor. Also given in the tables are the 95 percent Highest Posterior Density intervals (indicated by HPD lower and HPD upper) and the two-tailed MCMC probabilities p (for further details see Baayen et al. 2008).

In the next section we will present our analysis of words in accented contexts, in which we first investigate the acoustic parameters individually and then develop a model which can predict the stress pattern (left-prominent or right-prominent) on the basis of the acoustic parameters. The section to follow will then turn to unaccented contexts.

4 Words in accented position

4.1 Acoustic parameters

In this subsection we present the results of our mixed effects regression models for the five acoustic parameters. In these models we employed POSITION and STRESS as interacting predictors, and GENDER as a main effect with no further interactions to control for gender-specific differences. If one of the acoustic parameters is used differently for left- and right-prominent words, we expect to find a significant interaction between POSITION and STRESS for that parameter.

Overall, we found this crucial interaction of stress and position for three of the five parameters, i.e. pitch, intensity and spectral balance. In contrast, the measurements for duration and pitch slope did not depend on stress placement. Figure 1 illustrates the effects of stress and position on the three said parameters by plotting the means for each parameter for female speakers as predicted by our final models. In what follows we will discuss in more detail the analytical procedure and results for each parameter in turn, starting with F0.

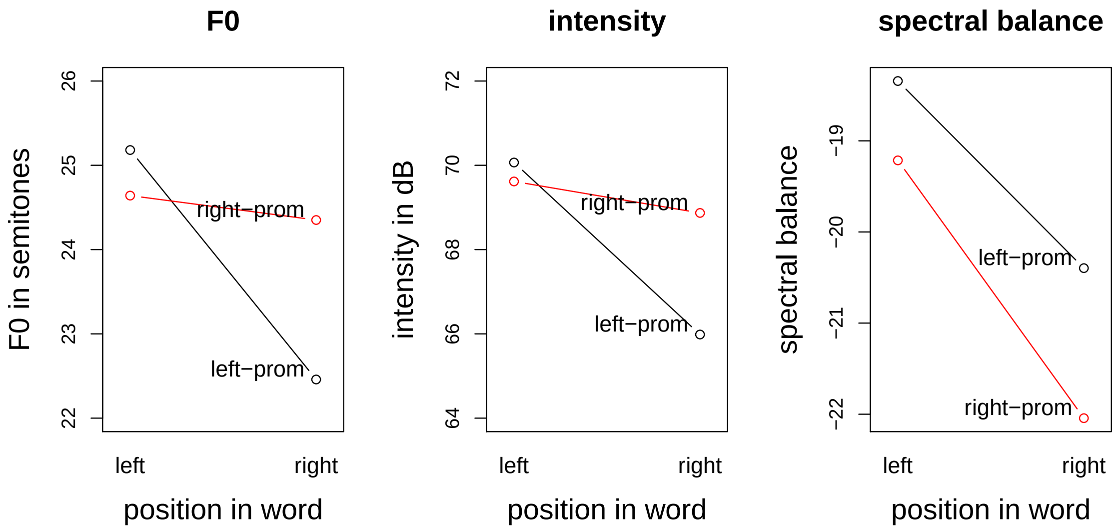

Figure 1: Interaction of prominence and position in the prediction of F0, intensity and spectral balance (accented words)

Since measurements were taken automatically, we unavoidably run the risk of obtaining sometimes extreme values that are simply measurement errors. The inspection of the distributions of the F0 measurements for male and female speakers showed that the data contained some rather extreme observations, which were clearly outside the ranges of normally distributed data. This concerned F0 measurements outside the range of 70 Hz to 150 Hz for male speakers, and 90 Hz to 230 Hz for female speakers. These extreme data points came from ten of the 19 different speakers (four males and six females). To avoid undue influence of these outliers on our models, measurements outside the above-mentioned ranges were excluded from the data, which resulted in an overall loss of 1.4 percent of the data, leaving 2354 observations (see Baayen and Milin 2010 for a discussion of data trimming procedures in the context of mixed effects regression analysis).

The model for F0, as given in Table 1, showed significant effects of position, stress and gender, and, crucially a significant two-way interaction of stress and position (p=0.0001). In the regression tables we give the names of variables in small capitals and the value of a variable in regular print, immediately following the variable name. The abbreviation 'right-prom' stands for 'right-prominent'.

Figure 1, left panel, shows the mean F0 values for female speakers as predicted by our final model. All values are about 6 semitones lower for male speakers, on average. We can see that right-prominent words are characterized by hardly any difference in left and right F0 (0.3 ST), whereas in left-prominent words, F0 in the right position is by 2.5 ST lower than in the left position. Also, F0 in the left position is significantly higher by 0.5 ST in left-prominent words than in right-prominent words.

For the intensity analysis, no prior trimming of the data was necessary. In analogy to the F0 analysis we devised a mixed-effects regression model with POSITION and STRESS as interacting predictors, and with GENDER as a main effect to control for possible differences in intensity between male and female speakers. The final model, as documented in Table

Table 1: Fixed-effect coefficients in a mixed-effects model fitted to F0 measurements in accented words, final model.

	Estimate	HPD lower	HPD upper	p
(Intercept)	25.1424	24.6853	25.6109	0.0001
POSITIONright	-2.4999	-2.6372	-2.3637	0.0001
STRESSright-prom	-0.5364	-0.7273	-0.3407	0.0001
GENDERmale	-6.0299	-6.7119	-5.3453	0.0001
POSITIONright:STRESSright-prom	2.2081	2.0202	2.3889	0.0001

2, showed a main effect for position and a significant two-way interaction of stress and position. Gender as a main effect was not significant.

Table 2: Fixed-effect coefficients in a mixed-effects model fitted to intensity measurements in accented words, final model.

	Estimate	HPD lower	HPD upper	p
(Intercept)	70.0666	69.1587	70.9280	0.0001
POSITIONright	-4.0815	-4.3654	-3.8132	0.0001
STRESSright-prom	-0.4498	-0.9146	-0.0119	0.0510
POSITIONright:STRESSright-prom	3.3345	2.9409	3.7388	0.0001

The mean intensity values as predicted by our final model are shown in Figure 1, middle panel. We can see that right-prominent words are characterized by a small positive difference between left and right intensity (0.7 dB, $p<0.001$), whereas left-prominent words have a very large positive difference (4.1 dB, $p<0.001$). Intensity in the left position is not statistically different between left-prominent and right-prominent words, but intensity drops notably for the right position of left-prominent words.

Let us turn to duration. The duration of a syllable largely depends on the vowel that forms the nucleus of the syllable. Thus, long vowels or diphthongs can be expected to show longer durations than short vowels, keeping other variables (such as position or stress) constant. Due to the phonological structure of our items, the vowels in the left position were mostly short vowels, while the vowels in right position were all diphthongs or long vowels. Henceforth, we will use 'long vowel' as a cover term for both long monophthongs and diphthongs.

The reason for this skewed distribution in our target words is that there are simply no right-prominent derived words in English with primary-stressed short vowels. Table 3 shows the distribution for the accented data set. Only the rows in bold print contain exclusively long vowels in both positions. We used this subset of 582 observations to model the effect of stress and position on duration in accented words.

Table 3: Distribution of vowels in left and right position, accented words

		Vright		
		aɪ	eɪ	iː
Vleft	ɔ	74	226	76
	ʌ	74	0	70
	æ	76	214	76
	aɪ	**72**	**224**	**64**
	ɛ	76	212	72
	ɚ	0	220	0
	ɪ	74	200	68
	uː	**0**	**220**	**0**

To control for the potential effect of different vowels (with potentially intrinsic differences in duration) within each position we added VLEFT and VRIGHT as covariates. Duration measurements were log-transformed to address concerns about curvature in the distribution of residuals (cf. Sluijter and van Heuven 1996a for the same procedure).

In our models, we do not find a significant effect for STRESS, nor for the crucial interaction between STRESS and POSITION. Thus, duration did not emerge as an acoustic correlate of the distinction between primary and secondary stress in accented words.

What about pitch slope? Prior to the analysis of pitch slope in accented words we inspected the distribution of pitch slope measurements and first removed 8 data points with extreme measurements (slopes beyond a span of -610 to 610, 0.3 percent of the observations), the resulting data set comprised 2382 observations. Pitch slope values ranged between -583.1 and 597.4 with half of the data ranging between -29.05 and 20.73. We fitted a mixed effects regression model with position, stress and gender as predictors to the accented data set, but no significant effects emerged.

For the analysis of spectral balance we fitted a mixed effects regression model with STRESS, POSITION and GENDER as predictors, with GENDER as a main effect. The final model showed a main effect for POSITION, and GENDER, and the crucial interaction of STRESS and POSITION. The model is documented in Table 4.

Table 4: Fixed-effect coefficients in a mixed-effects model fitted to measurements of spectral balance in accented words, final model.

	Estimate	HPD lower	HPD upper	p
(Intercept)	-18.3435	-19.398	-17.2250	0.0001
POSITIONright	-2.0532	-2.483	-1.6417	0.0001
STRESSright-prom	-0.8707	-1.844	0.1566	0.0872
GENDERmale	-1.5274	-2.905	-0.2322	0.0272
POSITIONright:STRESSright-prom	-0.7758	-1.363	-0.1755	0.0112

The right panel of Figure 1 shows the estimated means for accented words for females, the values for the males are about 1.5 units lower. We can see that left-prominent words have higher values than right-prominent words. In general, there is a higher spectral balance in the left position than in the right position (2.1 dB difference for left-prominent words, and 2.8 dB difference for right-prominent words), which means that in left positions, the distribution of energy in the spectrum is more balanced, i.e. has a less steep reduction of energy in the higher frequency band, than in the right position. Furthermore, spectral balance is higher for left-prominent words than for right-prominent words. It is particularly low for the right position of right-prominent words, i.e. the syllable which is expected to be prominent in this type of words.

To summarize, words, the position (left or right) and type of stress (primary or secondary) influences three of the five parameters, namely F0, intensity, and spectral balance. In contrast, the measurements for duration and pitch slope did not depend on stress position. In left-prominent words, we find large differences in F0 and intensity between the two stressed syllables, while right-prominent words show only small differences in F0 and intensity between the two stressed syllables. With regard to spectral balance, left-prominent and right-prominent words both show larger values in the left position than in the right position, but in left-prominent words the two values are higher than their corresponding values in right-prominent words.

4.2 Predicting stress patterns on the basis of acoustic parameters

In this subsection we will see how well we can predict the stress pattern of accented words on the basis of the five acoustic parameters we inspected in the previous section. As this analysis involves an estimation of the prominence in left position relative to that in right position, we calculated for each target word the difference between F0, intensity, and duration in left position and the corresponding measurement in right position. A positive value for the three measures derived in this way, *dPitch* for F0, *dIntensity* for intensity, *dDuration* for duration, thus indicates that the respective measure is larger in left position than in right position, while a negative difference represents the reverse case.

Similar difference measures have been used in earlier studies that investigated the prominence relation between two syllables, words, or elements of a compound, for instance in Morton and Jassem (1965), Farnetani et al. (1988), and Plag (2006). In these studies, difference measures were found to co-vary strongly with the prominence distinctions that were the focus of the investigation, which suggests a corresponding approach for the present regression model. Calculation of differences, instead of using raw measurements from the two syllables, has the additional advantage of reducing the amount of extra-linguistic variation. *dpitch* addresses differences in average fundamental frequency between different speakers, for instance between men and women. *dint* eliminates volume differences present in different recordings, which may be due to inconsistent microphone distances, amplification variations, or speaker inconsistencies. *ddur* accounts, at least partially, for differences in speaking rate that depend either on the speaker or on the specific text.

Thus, these difference measures help to incorporate into the model variation of F0, intensity, and duration that is presumably not related to the stress distinction under investigation here. For pitch slope and spectral balance, there is no reason to assume that difference measures are similarly appropriate. On the contrary, calculating the difference between slope measurements taken from left and right position may be counterproductive, as such a difference measure runs risk of obscuring the important local characteristics of pitch contour and frequency spectrum that are expressed by pitch slope and spectral balance. Therefore, pitch slope and spectral balance were included as separate predictors for left and right position, respectively.

As a first step in our analysis we took the whole data set and inspected the distributions of the measurements. The only trimming that seemed necessary concerned extreme measurements for the two pitch slopes, which resulted in the loss of 1.1 percent of the data. The final data set comprised 1184 accented items. We started our statistical analyses with the fixed and random effects shown in Table 5.

Table 5: Fixed and random effects initially entering the analysis

fixed effects
F0 difference
intensity difference
duration difference
pitch slope in left position
pitch slope in right position
spectral balance in left position
spectral balance in right position
random effects
speaker
item
slope for pitch difference
slope for intensity difference
slope for duration difference
correlations of random effects

Pitch slope in right position was not significant in the initial model and was therefore removed as a predictor. The final model for the accented words is documented in Table 6, the effects of the acoustic parameters are illustrated in the partial effects plots shown in Figure 2. Log-likelihood tests and increased values of C showed that the inclusion of random slopes for F0, intensity and duration, and of the correlations of random slopes for intensity and duration was justified. The model algorithm did not converge if the correlations of the random slope for F0 was also included, the correlations were therefore removed from the model. The inclusion of ITEM as random effect led to the insignificance of all fixed effects. This is to be expected, since there is no within-type variation concerning the dependent variable STRESS. Hence, ITEM was dropped as a random effect.

Table 6: Fixed-effect coefficients, standard error, z-values and p-values in a generalized mixed-effects model fitted to the logits of right prominence in accented words, final model. C=0.9495955, Dxy=0.8991910

	Estimate	Std. Error	z-value	p
(Intercept)	-1.802807	0.787602	-2.289	0.022080
dPitch	-0.619430	0.133654	-4.635	3.58e-06
dIntensity	-0.286336	0.070769	-4.046	5.21e-05
dDuration	-8.834548	2.384485	-3.705	0.000211
leftBalance	-0.039555	0.018138	-2.181	0.029198
rightBalance	-0.158009	0.030517	-5.178	2.25e-07
leftSlope	-0.003562	0.001367	-2.605	0.009182

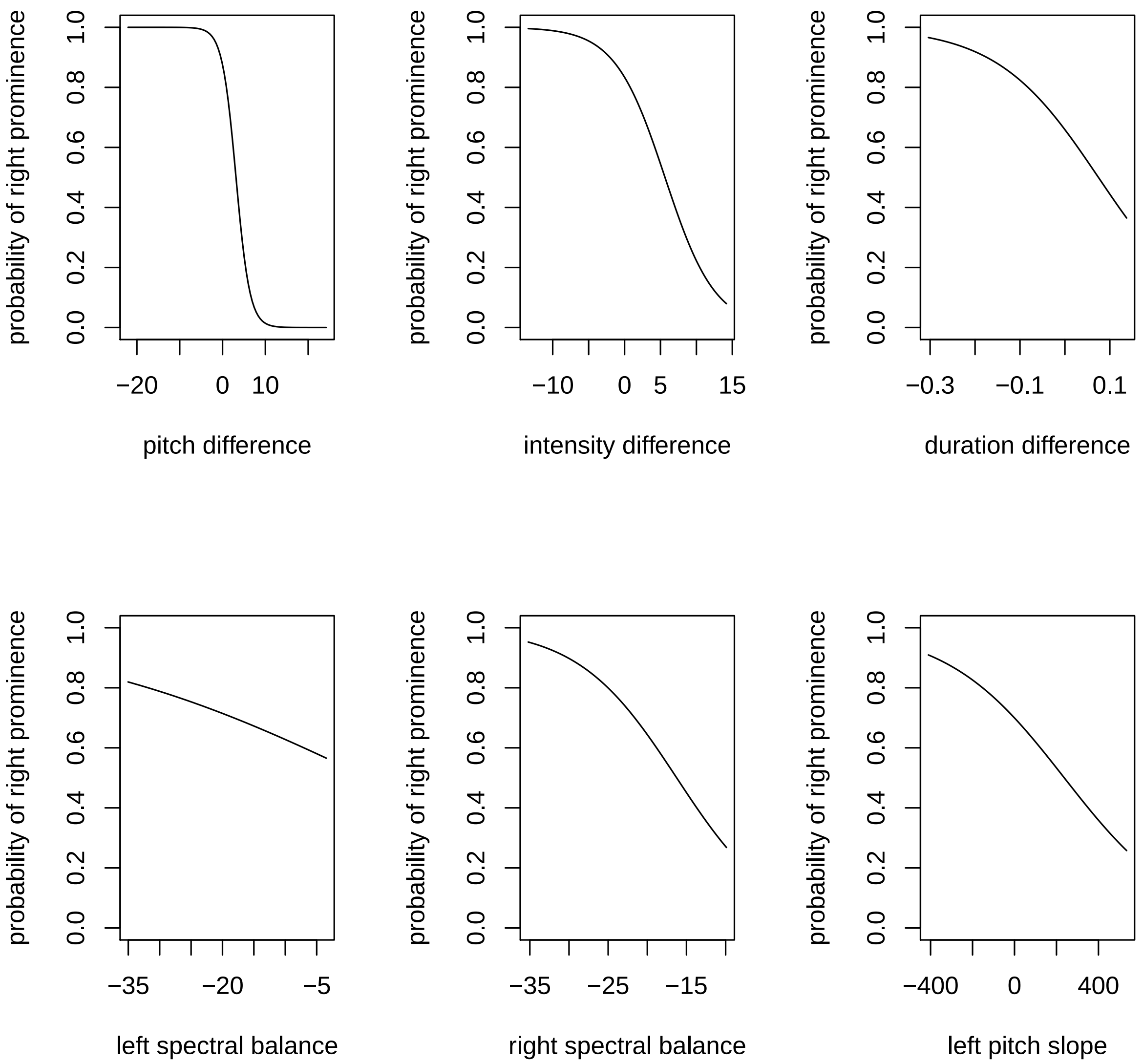

Figure 2: Partial effects for accented words

Both the regression coefficients and the partial effects plots show that the chances of right prominence increase with decreasing values for each of the six predictors. The strongest effect can be seen for *dPitch* and *dIntensity*, as the probability of right prominence changes drastically for the different observed values of these predictors. In comparison, the effects of durational differences and left spectral balance are clearly less distinctive. Thus, F0 and intensity appear to provide the strongest cues to left and right prominence. With a C-value of 0.95 the model is extremely successful in its predictions.

5 Words in unaccented position

5.1 Acoustic parameters

In this subsection we present the results of our regression analyses of the five acoustic parameters for words in unaccented position. These models were derived in an analogous fashion as those in section 4.1. We found the crucial interaction of STRESS and POSITION for the same three parameters as before, i.e. pitch, intensity and spectral balance. In contrast, the measurements for duration and pitch slope did not depend on stress placement. Figure 3 plots the significant interactions of prominence and position in the prediction of the three said parameters. The y-axes show the means for each parameter for female speakers as predicted by our final models.

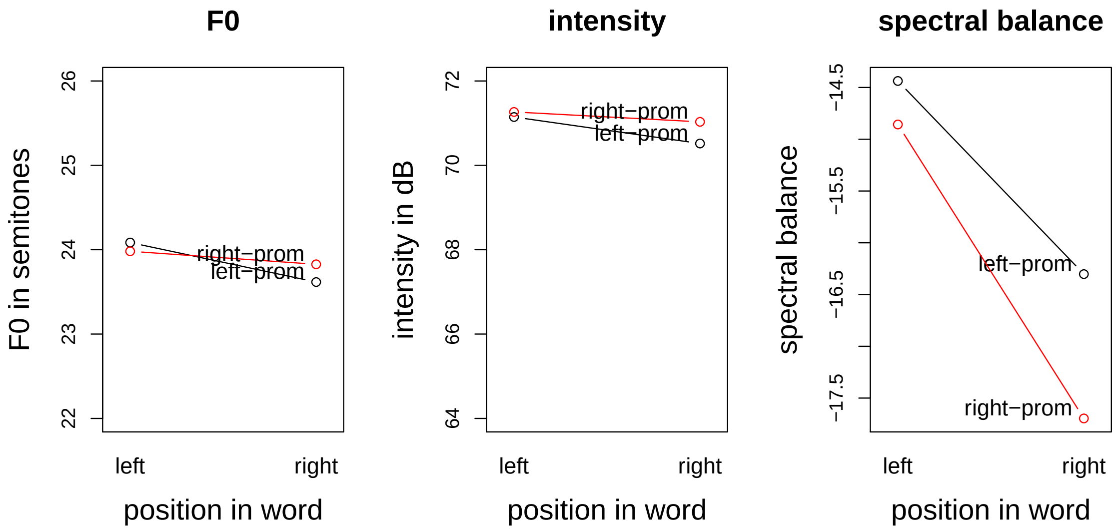

Figure 3: Interaction of prominence and position in the prediction of F0, intensity and spectral balance (accented words)

We find only small differences in F0 and intensity between different positions and different prominence patterns. In other words, the effect sizes are rather small. Spectral balance shows a different picture, in that left-prominent and right-prominent words both show larger values in the left position, but left-prominent words show higher values than right-prominent words. We will now discuss each parameter in more detail.

Problems of extreme F0 measurements occurred also with unaccented words (as described for the accented data above), most of which were cases of apparently erroneous measurements. An inspection of the distributions of the measurements showed that values between 51 Hz and 120 Hz (male speakers) and between 90 Hz and 210 Hz (female speakers) were plausible measurements, and data points ouside that range were removed (i.e. 4.3 percent of the data), reducing the data set to 2550 observations.

The model for unaccented words, as given in Table 7, showed main effects of POSITION and GENDER and a significant interaction of STRESS and POSITION.

Table 7: Fixed-effect coefficients in a mixed-effects model fitted to F0 measurements in unaccented words.

	Estimate	HPD lower	HPD upper	p
(Intercept)	24.0847	23.7082	24.4893	0.0001
POSITIONright	-0.4681	-0.5609	-0.3747	0.0001
STRESSright-prom	-0.1032	-0.2271	0.0150	0.0924
GENDERmale	-8.0340	-8.6622	-7.4184	0.0001
POSITIONright:STRESSright-prom	0.3137	0.1832	0.4438	0.0001

The left panel of Figure 3 shows the mean values of F0 for female speakers; the values for male speakers are about 8 semitones lower. Alternation of the contrast codings for the factors POSITION and STRESS show that F0 is not significantly different in the left positions of left-prominent and right-prominent words (p=0.09). However, the change from the left to the right position in the two types of words is significantly different. The difference in right-prominent words amounts to only 0.15 ST (p<0.01), while it is 0.46 ST for left-prominent words (p<0.01). Accordingly, F0 in the right position is significantly lower in left-prominent words than in right-prominent words (0.21 ST, p<0.01)

Given these small differences, it seems that left-prominent words and right-prominent words are not well discriminated by F0 if they are unaccented. This impression is corroborated by the fact that the model for unaccented words has increasingly larger residuals for more extreme measurements, which shows that in unaccented words the F0 of the stressed syllables is not so well predictable as with accented words. Together with the increased imprecision of the regression model (in comparison to that for accented words), the results for unaccented words can be interpreted in such a way that F0 is not a very good indicator of the distinction between primary and secondary stress in unaccented words.

Let us turn to intensity. The raw data contained some observations with extremely low intensity (I<60 dB) which, in view of the overall distribution of the data, could be considered as clear outliers. The removal of these observations resulted in the loss of 0.7 percent of the observations. In the model for the unaccented words (8) we find a significant main effect of position and a significant interaction of position and stress.

The mean intensity values for unaccented words as predicted by our model are shown in the middle panel of Figure 3. Right-prominent words are characterized by a very small positive difference in left and right intensity (0.2 dB). Left-prominent words show a larger difference of the same kind (0.6 dB). Intensity in the left position is not statistically different between left-prominent and right-prominent words, and drops slightly, but

Table 8: Fixed-effect coefficients in a mixed-effects model fitted to intensity measurements in unaccented words.

	Estimate	HPD lower	HPD upper	p
(Intercept)	71.1456	70.2968	72.0049	0.0001
POSITIONright	-0.6280	-0.8800	-0.3585	0.0001
STRESSright-prom	0.1202	-0.2951	0.5479	0.5828
POSITIONright:STRESSright-prom	0.3930	0.0254	0.7619	0.0358

significantly, for the right position of left-prominent words.

For the analysis of duration for the unaccented words we again took only those words into consideration that had long vowels in both positions (N = 656). We fitted a linear mixed-effects model with POSITION and STRESS as interacting predictors, GENDER as main effect, and VLEFT and VRIGHT as covariates. Apart from the effects of the control variables VLEFT and VRIGHT we find no significant effects. In particular, the interaction of POSITION and STRESS reaches only a p-value of 0.11.

For pitch slope, we again fitted a mixed effects regresssion model with POSITION, STRESS and GENDER as predictors, but no significant effects emerged.

In the analysis of spectral balance the model (documented in Table 9) shows a main effect of POSITION and an interaction of STRESS and POSITION.

Table 9: Fixed-effect coefficients in a mixed-effects model fitted to measurements of spectral balance in unaccented words, final model.

	Estimate	HPD lower	HPD upper	p
(Intercept)	-14.4372	-15.722	-13.2092	0.0001
POSITIONright	-1.8654	-2.265	-1.4874	0.0001
STRESSright-prom	-0.4202	-1.375	0.5521	0.3996
POSITIONright:STRESSright-prom	-0.9738	-1.525	-0.4391	0.0002

In the right panel of Figure 3 we find that, similar to accented words, the right position of unaccented words has a significantly lower value for spectral balance than the left position (1.9 dB difference for left-prominent words, and 2.8 dB difference for right-prominent words). Spectral balance is particularly low for the right position of right-prominent words, while there is no statistical difference between the left positions of left-and right-prominent words.

To summarize the discussion of the acoustic parameters for unaccented words we can say that the position (left or right) and type of stress (primary or secondary) significantly influences F0, intensity, and spectral balance, but not duration and pitch slope. The effects for F0 and intensity are, however, quite small.

5.2 Predicting stress patterns on the basis of acoustic parameters

To test whether stress patterns of unaccented words are predictable on the basis of the acoustics, we fitted a mixed-effects regression model with the same predictors as given in Table 5 above. The final model is documented in Table 10. Log-likelihood tests showed that random slopes were not justified, the model has only SPEAKER as a random effect. Only F0 difference and spectral balance in the right position turned out to be significant predictors. The model's predictive power is very moderate ($C = 0.60$), which is also evident from the partial effects plots shown in Figure 4. The plots show that even the most extreme values of F0 and right spectral balance do not lead to very high or very low probabilities, i.e. to clear decisions. For the range of F0 differences between -5 and 5 ST, which includes about 90 percent of all F0 differences, the probability for either prominence pattern changes by about 0.15, so for the vast majority of observations, the F0 difference cannot be considered a reliable predictor of right-prominence. This is reflected in the low C-value of the model.

Table 10: Fixed-effect coefficients, standard error, z-values and p-values in a generalized mixed-effects model fitted to the logits of right prominence in unaccented words. C=0.595039, Dxy=0.190078

	Estimate	Std. Error	z-value	p
(Intercept)	-0.97796	0.20982	-4.661	3.15e-06
DPITCH	-0.03650	0.01235	-2.955	0.00313
RIGHTBALANCE	-0.05885	0.01188	-4.956	7.21e-07

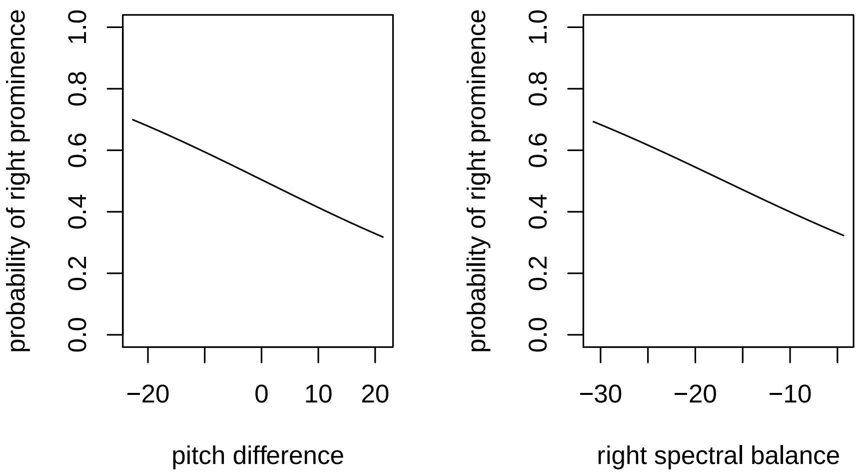

Figure 4: Partial effects for unaccented words

Intensity, duration, and pitch slopes do not come out as significant correlates of primary and secondary stress in unaccented words, which corroborates our findings from the previous subsection that primary and secondary stress in unaccented words is not well distinguishable on the basis of these acoustic cues.

6 A comparison of accented and unaccented words

In this section, we compare the results for accented and unaccented words. Table 11 summarizes our results for the acoustic parameters.

Table 11: Correlates of prominence

	accented words	**unaccented words**
F0	yes, strong	yes, very weak
intensity	yes, strong	yes, very weak
duration	no	no
pitch slope	no	no
spectral balance	yes	yes

We can see that accented and unaccented words differ in how far the five parameters are affected by primary and secondary stress placement. While for accented words we find strong effects of stress position and prominence-pattern on F0 and intensity, unaccented words show only weak effects. Spectral balance is affected by stress position and prominence pattern in both accented and unaccented words.

The fact that the accented and unaccented data sets originate from two different groups of speakers raises, however, some questions about the interpretation of these differences and similarities. One potential confound, for instance, is the possibility of dialectal differences between the groups. It is not unthinkable that varieties of English may differ in the extent to which they use acoustic cues to signal differences in stress level. Kochanski et al. (2005), for instance, find that F0 as a cue to accentuation varies across some British accents. Studies such as Clopper and Pisoni (2004) show that American English dialects differ with regard to certain properties of certain vowels and consonants, but relatively little is known about how American English dialects may differ in their prosody.

It is, however, generally assumed that English in North America constitutes a much more homogeneous variety than English on the British Isles. Canada in particular has very little dialectal diversity within its boundaries, and the phonological and phonetic differences between the varieties spoken in Canada and in Western and Midland Regions of the United States are, with notable exceptions such as Canadian Raising, also rather small, so that it seems fairly safe to compare data obtained from populations in Canada and California (e.g. Trudgill and Hannah 2002, Labov et al. 2006).

Another problem arises through the use of different recording equipment and different recording environments. These factors may especially influence measurements of intensity and spectral balance (the latter for example due to different frequency responses of the microphones used).

Finally, the two conditions are not fully comparable for F0. In the unaccented condition the items we are interested in are in a non-final intonational phrase, while in the accented condition the items are in a final intonational phrase (see again section 3.1 above). Non-final intonational phrases are expected to be higher in pitch than final ones.

The problems arising from differences in the recording environment, the recording equipment and the position of the item's intonational phrase would forbid a meaningful comparison of absolute measurements across data sets. For example, to compare the mean F0 values of accented left-prominent words with the mean F0 values of unaccented left-prominent words would not be very helpful since the different values come from different recording environments and equipments. It seems, however, justified, to compare the relationship between two positions in the accented data set with the same relationship between the two positions in the unaccented data set.

If we do that and compare figures 1 and 3, we can see the following interesting differences between the two conditions. While left-prominent accented words show a large difference in F0 and intensity between the two stressed positions, left-prominent unaccented words show more or less level F0 and intensity. Right-prominent accented words show only a slight decrease in F0 and intensity, and right-prominent unaccented words show level F0 and intensity from left to right position. The patterns for spectral balance are very similar across the two data sets.

7 Discussion

7.1 Acoustic correlates of primary and secondary stress

We investigated in this paper whether some acoustic properties of stressed syllables (F0, intensity, duration, pitch slope, and spectral balance) depend significantly on the position of primary and secondary stress within the word, and on whether the word is accented or not. It was shown that the position within the word and the presence or absence of accentuation significantly influence F0, intensity, and spectral balance, but not duration or pitch slope of the syllable in question.

Not surprisingly, we found differences between accented words and unaccented words concerning the strength of the three significant parameters. In accented, left-prominent words, the differences in F0 and intensity between the two stressed syllables are large, while unaccented left-prominent words show much smaller (though still significant) differences in F0 and intensity. Right-prominent words generally show smaller differences in F0 and intensity between the two stressed syllables, and these differences are even less pronounced for unaccented words. In contrast, accented and unaccented words do not differ significantly in their relation of spectral balance and the two stressed positions.

In a conceptually different analysis, we tried to predict the stress pattern of a given word based on the same five acoustic parameters as in the previous analyses. In this analysis, accented words behave very differently from unaccented words. For accented words, all five parameters turned out to be predictive for the stress pattern, and it was possible to provide a highly successful statistical model to detect the correct prominence pattern. This was different for unaccented words, where only F0 and spectral balance are significant predictors, and a model based on these parameters is not able to successfully distinguish unaccented left-prominent from unaccented right-prominent words.

It is not easy to relate our findings to the results of previous studies, since these studies mostly investigated the contrasts between stressed and unstressed syllables and/or did not take accentuation into account. With regard to the phonetic correlates of primary and secondary stress, our findings are nevertheless largely compatible with that literature.

For example, Sluijter & van Heuven (1996b) found that pitch and intensity cue stress, but only in accented words. They looked at the stressed-unstressed distinction, but their results largely parallel ours for the primary-secondary stress distinction in that we find F0 and intensity cueing the primary-secondary stress distinction in accented words. Mattys (2000) looked at three different contrasts in only accented words: primary stress vs. unstressed, secondary stress vs. unstressed, and primary vs. secondary stress in the left position across words. His results are similar to ours in that he finds that listeners' discrimination of the two stresses in accented words depended on pitch, intensity and duration. Note, however, that in our data we only find a duration effect in one type of analysis, i.e. in the prediction of stress for accented words.

Fear et al. (1995) only find a duration effect but not one of either intensity, pitch or spectral characteristics when comparing primary stress with secondary stress vowels, but listeners' judgments could be predicted on the basis of intensity and spectral measurements. Pitch did not turn out to be significant in that experiment, but the authors compared primary and secondary stress vowels in the left position across words, and not within words. This means that their results are very much in accordance with our own. We only found a small difference in F0 and intensity between the left position of right-prominent word and the left position of left-prominent words (see again Figure 1, and the pertinent discussion in section 4.1).

Our findings are also compatible with Oboki (2006), who also found that accented and unaccented words differ in their realization of stress. Further similarities between Oboki's and our study include the very important role of F0 and intensity with accented words, and, with unaccented words, the significance of spectral balance. The effect of spectral balance is, however, not uncontroversial. Sluijter & Heuven (1996a), for instance, find that a stressed syllable has an increased intensity in the higher band of the spectrum when compared to an unstressed syllable. Since, in our calculation of spectral balance, we subtract intensity in the low-frequency band from intensity in the high-frequency band ($B = I_{high} - I_{low}$, cf. section 3.2), an increase of intensity in the high-frequency band I_{high} would result in a higher spectral balance. Translating the findings by Sluijter and van Heuven (1996b) accordingly, we would expect stressed syllables to have a higher spectral balance than unstressed syllables. Again, unlike Sluijter and van Heuven, we are not dealing with a stressed vs. unstressed opposition in the present paper, but still it is very unexpected to find that spectral balance is generally higher in the left than in the right position, regardless of the prominence pattern. In particular, it remains unclear why spectral balance should be higher in the non-prominent (right) position of left-prominent words than in the prominent (right) position of right-prominent words.

The role of duration is somewhat surprising, as duration doesn't seem to differ between syllables with primary stress and secondary stress neither in accented nor in unaccented words. Durational differences have been found in other studies (e.g. Fear et al. 1995b, Mattys 2000b, de Jong 2004) to be related to the difference between primary and secondary stress, with primarily stressed syllables being generally longer (see also Sluijter and van Heuven 1996a).

One possible explanation for the lack of a durational difference between the syllables in our left-prominent accented words may lie in the observation reported by Turk and White (1999). They find that the lengthening effect of a pitch accent is not restricted to the syllable it is associated with, but seems to spread throughout the whole word. Thus, a potential duration difference between the left (accented) and the right (unaccented) syllable in left-prominent words might be obscured by such a durational spread. However, it remains unclear why this phenomenon might have such an obscuring effect in the present data, but not in Fear et al. (1995b) or Mattys (2000b).

Given that no durational difference between the left and the right position could be identified in any of our conditions, we may (perhaps somewhat hesitantly) conclude that duration may be a strong acoustic correlate which separates unstressed from stressed syllables, but which does not contribute very much to signal the difference between syllables with primary or secondary stress, or accented and unaccented (but stressed) syllables in a word. Thus, although duration is one of the two phonetic correlates that Gussenhoven (2004:22) proposes as possibly distinguishing primary and secondary stress in English (the other being related to the timing of pitch movements), the evidence for such a distinction does not seem to be very strong. Spitzer et al. (2007) produce partial independent support for this conclusion, as they find that listeners do not depend strongly on length differences in stress-based segmentation tasks.

Overall, our analysis has substantiated the role of F0, intensity, pitch slope, and spectral balance as correlates of the distinction between primary and secondary stress in accented words, largely in accordance with related findings in the literature. The role of duration remains somewhat unclear, however. For unaccented words, we have discovered that the distinction between the two stresses becomes largely neutralized.

7.2 Phonological interpretation

Our results may have important implications for an understanding of the nature of primary and secondary stress in phonology. For accented words, we found clear differences of F0 and intensity between left-prominence and right-prominence, while in unaccented words there are only slight differences of F0 and intensity. Irrespective of accentuation, there is a lower spectral balance in right position for right-prominent words. How can these findings be interpreted phonologically? Let us first focus on accented words.

In left-prominent accented words, there is a steep drop of F0 and intensity from left to right position; the right position of these words is very markedly different from the left position. As high F0, but also high intensity, have been associated with accented syllables, it seems safe to assume that the left position of left-prominent accented words is accented, while the right position of these words does not carry an accent.

But what about right-prominent accented words? First of all, there is no large difference between the left and right position in these words. In addition, the left position is acoustically very similar to that of left-prominent accented words (which we have just concluded to be accented), while the right position differs strongly from the same position in left-prominent accented words (which is apparently unaccented). Thus, we conclude that in right-prominent accented words there is an accent on both positions, which is fully compatible with the assumptions made in the metrical phonology literature as summarized above in section 2.

In sum, our data for accented words strongly indicate that the difference between left-prominent and right-prominent words is realized by accent placement. Left-prominent words have a single accent on the first strong syllable (counting from the left), right-prominent words have two accents, i.e. one on each of the two strong syllables. This implies that the notion of secondary stress is to some extent problematic. While the primary stress syllable and the secondary stress syllable are both strong syllables irrespective of their respective positions within the word, the acoustics as well as the phonology (in terms of accentuation) of the two secondary stress syllables in right-prominent vs. left-prominent words are quite different from each other.

When comparing the left syllables of accented left- and right-prominent words, there is a significant F0 difference of about 0.5 ST. Hence, the accentuation of the left syllable already provides an acoustic cue whether the whole word will be left-prominent or right-prominent. This explains why the participants in Mattys (2000) could distinguish primary and secondary stressed initial syllables, and it is also something that was found for the left elements of compounds (Kunter 2010). As mentioned above, Fear et al. (1995:1896) find no significant difference of the acoustic correlates (apart from duration) between primary stress and secondary stress in what we consider the left position of accented words. This agrees very well with our data for F0 and intensity, and all this is very similar to what Kunter found for compounds. As a consequence, one could even claim that, phonologically, the difference between primary and secondary stress in accented words is the same as that in compounds. The success of the predictive model in section 4.2 indicates that the distinction between left-prominence and right-prominence in accented words is very strong, and can easily be predicted from the acoustic correlates.

In sum, it seems that there is not much evidence for assuming that English features a phonetic distinction between primary and secondary stress; the three-way distinction outlined above (unstressed – stressed unaccented – stressed accented) seems to be sufficient to account for the acoustic data in our experiments. Of course, *isolate* and *isolation* have different prominence patterns, but these patterns are distinguished by a difference in the distribution of accents, and thus, this difference emerges only if the words occur in accented position.

Let us turn to unaccented words. We do find a drop of F0 and intensity from left to right in left-prominent unaccented words, too, but this drop is much smaller, and perceptually probably much less salient than in accented words. It is unclear whether the differences are still sufficiently large to be perceived as two distinct levels of stress. The perception studies by Mattys (2000) and Fear et al. (1995) do not help us here, as they are based on accented words. Thus, future studies will have to find out whether this acoustic difference is large enough to be cue perceptually a difference in prominence.

The observation that F0 and intensity in the right position are always lower than in the left position, irrespective of accentuation or prominence, can straightforwardly be attributed to the declination effect described, for instance, by Collier (1975): due to the steady decrease of subglottal pressure, F0 and intensity are expected to decrease steadily throughout an utterance, disregarding any influence of the overall intonation pattern. In particular, this explains the similarities between right-prominent words in both the accented and the unaccented condition. The peak heights of the two subsequent accents in the former case are similarly affected by an overall declination as the unaccented subsequent syllables in the latter case. One reviewer pointed out that in principle, a slightly lower F0 in the right position of right-prominent words could potentially also be analyzed as a downstepped pitch accent, e.g. !H* in ToBI notation. Yet, given that the pitch difference between left and right syllable is very similar in three of the four conditions – accented right-prominent, unaccented left-prominent, and unaccented right-prominent –, it seems plausible to assume the same effect for all of them. And a downstepped pitch accent in the unaccented conditions seems rather unlikely.

Yet, it is noteworthy that there is a small, but significant difference in the first position between primary and secondary stressed syllables that can also be perceived by speakers and used as a cue to prominence interpretation (as shown by Fear et al. 1995b and Mattys 2000b). If this is so, the distinction between primary and secondary stress has a reality, at least under accentuation. Obviously, the realization of a pre-nuclear accent (i.e. the left accent in a right-prominent word, in which the nuclear accent falls on the right syllable) tends to be different, and thus, provide different acoustic cues, to that of a nuclear accent (i.e. the left accent in a left-prominent word). Such differences between pre-nuclear and nuclear accents have been described for instance in Silverman et al. (1990) who focus on peak alignment in these different conditions. Alignment was not taken into consideration in the present study, but it might be that it is this acoustic difference that contributes to the perceptual difference.

Returning to unaccented words, differences between the two positions are only very minor, and the prominence pattern cannot be predicted well from the acoustic parameters. It is unclear yet whether listeners can perceive a prominence difference between left and right position in unaccented words, after all. The acoustic cues seem to be so weak that it may not be possible.

According to the above reasoning, primary and secondary stress syllables are stressed syllables that are different from unstressed syllables, but not from each other, unless the word is accented. In this case the target of a nuclear accent corresponds to what is usually labeled the primary stress syllable. If the word occurs in an environment in which no accents are present (e.g. in post-nuclear position), there is no phonological difference between the first and third syllable in words such as *i.so.late* and *i.so.la.tion*. These syllables are simply stressed (or strong) syllables (which, of course, differentiates them from the second syllable *.so.*, which is unstressed, or weak).

To summarize, based on our findings, we claim that the acoustic evidence points towards a phonological account according to which, if we disregard accentuation, there is no difference between secondary and primary stress. Instead, in accented words, the prominence difference between *isolate* and *isolation* should be considered a difference of presence or absence of an accent on the third syllable. Finally, we claim that in unaccented words, the prominence pattern of *isolate* and *isolation* is basically the same: S-w-S and S-w-S-w, with no phonological and probably also no perceptible acoustic difference between primary and secondary stress.

References

Baayen, R. H. and Milin, P. (2010). Analyzing reaction times. *International Journal of Psychological Research*.

Bates, D., Sarkar, D., Bates, M., and Matrix, L. (2007). The lme4 package. *R Package Version 2.4*, 1.

Beckman, M. E. (1986). *Stress and non-stress accent*. Foris Publications, Dordrecht.

Beckman, M. E. and Edwards, J. (1994). Articulatory evidence for differentiating stress categories. In Keating, P., editor, *Phonological structure and phonetic form*, Papers in Laboratory Phonology III, pages 7–33. Cambridge University Press, Cambridge.

Boersma, P. and Weenink, D. (2010). Praat: doing phonetics by computer. *computer program, version 4.5.16, http://www. praat. org/*.

Campbell, N. and Beckman, M. E. (1997). Stress, prominence, and spectral tilt. In Botinis, A., Kouroupetroglou, G., and Carayiannis, G., editors, *Proceedings of the ESCA Workshop Intonation: Theory, Models and Applications, September 18-20*, pages 67–71, Athens.

Clopper, C. and Pisoni, D. (2004). Some acoustic cues for the perceptual categorization of American English regional dialects. *Journal of Phonetics*, 32(1):111–140.

Collier, R. (1975). Physiological correlates of intonation patterns. *Journal of the Acoustical Society of America*, 58:249–255.

de Jong, K. (2004). Stress, lexical focus, and segmental focus in English: patterns of variation in vowel duration. *Journal of Phonetics*, 32:493–516.

de Jong, K., Beckman, M. E., and Edwards, J. (1993). The interplay between prosodic structure and coarticulation. *Language and Speech*, 36(2, 3):197–212.

Farnetani, E., Torsello, C. T., and Cosi, P. (1988). English compound versus non-compound noun phrases in discourse: an acoustic and perceptual study. *Language and Speech*, 31(2):157–180.

Fear, B., Cutler, A., and Butterfield, S. (1995a). The strong/weak syllable distinction in English. *Journal of the Acoustical Society of America*, 97(3):1893–1904.

Fear, B. D., Cutler, A., and Butterfield, S. (1995b). The strong/weak syllable distinction in English. *Journal of the Acoustical Society of America*, 97(3):1893–1904.

Fry, D. (1958). Experiments in the perception of stress. *Language and speech*, 1(2):126–152.

Fry, D. B. (1955). Duration and intensity as physical correlates of linguistic stress. *Journal of the Acoustical Society of America*, 32:765–9.

Giegerich, H. (1992). *English phonology: An introduction*. Cambridge Univ Press.

Gussenhoven, C. (2004). *The phonology of tone and intonation*. Cambridge University Press, Cambridge.

Hammond, M. (1999). *The phonology of English: a prosodic optimality-theoretic approach*. Oxford University Press, USA.

Harrington, J., Beckman, M. E., Fletcher, J., and Palethorpe, S. (1998). An electropalatographic, kinematic, and acoustic analysis of supralaryngeal correlates of word and utterance-level prominence contrasts in English. In *Proceedings of the 1998 International Conference on Spoken Language Processing*, volume 1854.

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

Hermes, D. J. and 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.

Heuven, V. J. v. (1987). Stress patterns in Dutch (compound) adjectives: acoustic measurements and perception data. *Phonetica*, 44:1–12.

Hirschberg, J. (2004). Pragmatics and intonation. In Horn, L. R., editor, *Handbook of pragmatics*, pages 515–537. Blackwell, Malden, MA.

Johnson, K. (2003). *Acoustic and auditory phonetics*. Blackwell, Malden, Mass., 2nd edition.

Kochanski, G., Grabe, E., Coleman, J., and Rosner, B. (2005). Loudness predicts prominence: fundamental frequency lends little. *Journal of the Acoustical Society of America*, 118(2):1038–1054.

Kunter, G. (2010). *Compound stress in English: The phonetics and phonology of prosodic prominence*. Niemeyer, Tuebingen.

Kunter, G. and Plag, I. (2007). What is compound stress? In *Proceedings of the 16th International Congress of Phonetic Sciences, Saarbrücken*, pages 1005–1008.

Labov, W., Ash, S., and Boberg, C. (2006). *The atlas of North American English: phonetics, phonology, and sound change: a multimedia reference tool*. Walter De Gruyter, Berlin/New York.

Laver, J. (1994). *Principles of phonetics*. Cambridge University Press, Cambridge.

Lieberman, P. (1960). Some acoustic correlates of word stress in American English. *Journal of the Acoustical Society of America*, 32(4):451–454.

Mattys, S. (2000a). The perception of primary and secondary stress in English. *Perception & psychophysics*, 62(2):253–265.

Mattys, S. L. (2000b). The perception of primary and secondary stress in English. *Perception & Psychophysics*, 62(2):253–265.

Morton, J. and Jassem, W. (1965). Acoustic correlates of stress. *Language and Speech*, 8:159–181.

Oboki, A. (2006). *Acoustic correlates of word stress in American English*. PhD thesis, MIT.

Odden, D. (2005). *Introducing phonology.* Cambridge university press.

Pierrehumbert, J. B. (1980). *The phonology and phonetics of English intonation.* PhD thesis, Massachusetts Institute of Technology.

Plag, I. (2006). The variability of compound stress in English: structural, semantic, and analogical factors. *English Language and Linguistics*, 10(1):143–172.

Silverman, K., , and Pierrehumbert, J. B. (1990). The timing of prenuclear high accents in English. In Kingston, J. and Beckman, M. E., editors, *Between the Grammar and Physics of Speech*, Papers in Laboratory Phonology I. Cambridge University Press, Cambridge.

Sluijter, A. M. and van Heuven, V. J. (1996a). Acoustic correlates of linguistic stress and accent in Dutch and American English. In *Proceedings of the 4th International Conference on Spoken Language Processing, Philadelphia*, volume 3, pages 630–633.

Sluijter, A. M. and van Heuven, V. J. (1996b). Spectral balance as an acoustic correlate of linguistic stress. *Journal of the Acoustical Society of America*, 100(4):2471–2485.

Spitzer, S. M., Liss, J. M., and Mattys, S. L. (2007). Acoustic cues to lexical segmentation: A study of resynthesized speech. *Journal of the Acoustical Society of America*, 122(6):3678–3687.

Terken, J. and Hermes, D. J. (2000). The perception of prosodic prominence. In Horne, M., editor, *Prosody: theory and experiment. Studies presented to Gösta Bruce.* Kluwer Academic Publishers, Dordrecht.

Trudgill, P. and Hannah, J. (2002). *International English. A guide to the varieties of Standard English.* Arnold, London.

Turk, A. E. and White, L. (1999). Structural influences on accentual lengthening in English. *Journal of Phonetics*, 27:171–206.

Štekauer, P., Zimmermann, J., and Gregová, R. (2007). Stress in compounds: an experimental research. *Acta Linguistica Hungarica*, 54(2):193–215.

8 Appendix: Items

Table 12: Target words

left-prominent	**right-prominent**
illuminate	illumination
elucidate	elucidation
hallucinate	hallucination
indicate	indication
imitate	imitation
irritate	irritation
abdicate	abdication
activate	activation
calculate	calculation
compensate	compensation
concentrate	concentration
conjugate	conjugation
delegate	delegation
escalate	escalation
emulate	emulation
hibernate	hibernation
violate	violation
isolate	isolation
terminate	termination
circulate	circulation
perforate	perforation
randomize	guarantee
dramatize	manatee
terrorize	refugee
temporize	legatee
modernize	nominee
solemnize	promisee
stigmatize	internee
victimize	disagree
publicize	publishee
customize	governee
privatize	licensee
finalize	financee