

Signal-based and expectation-based factors in the perception of prosodic prominence

JENNIFER COLE, YOONSOOK MO, MARK HASEGAWA-JOHNSON

University of Illinois

Abstract

The perception of prosodic prominence in spontaneous speech is investigated through an online task of prosody transcription using untrained listeners. Prominence is indexed through a probabilistic prominence score assigned to each word based on the proportion of transcribers who perceived the word as prominent. Correlation and regression analyses between perceived prominence, acoustic measures and measures of a word's information status are conducted to test three hypotheses: (i) prominence perception is signal-driven, influenced by acoustic factors reflecting speakers' productions; (ii) perception is expectation-driven, influenced by the listener's prior experience of word frequency and repetition; (iii) any observed influence of word frequency on perceived prominence is mediated through the acoustic signal. Results show correlates of perceived prominence in acoustic measures, in word log-frequency and in the repetition index of a word, consistent with both signal-driven and expectation-driven hypotheses of prominence perception. But the acoustic correlates of perceived prominence differ somewhat from the correlates of word frequency, suggesting an independent effect of frequency on prominence perception. A speech processing account is offered as a model of signal-driven and expectation-driven effects on prominence perception, where prominence ratings are a function of the ease of lexical processing, as measured through the activation levels of lexical and sub-lexical units.

1. Introduction

The spoken form of a word depends on two main factors: the vowels and consonants that make up the word, and the prosodic context, which includes the syllable structure, metrical structure at the word and phrase levels, and phonological phrase structure. This paper is concerned with the production and perception of prosodic context in ordinary speech, and specifically with prosodic *prominence*, or the strength of a spoken word relative to the words surrounding it in the utterance.

Prominence derives from metrical structure as an aspect of phonological representation (Beckman 1996; Ladd 1996). A syllable or word is prominent if it is parsed in a strong position in metrical structure. In languages with word-level stress, such as English, the parsing of syllables into metrical feet within the word is highly constrained by phonological parameters that govern, e.g., the location of strong syllables relative to the word edge. Let us refer to prominence in this sense as “structural” prominence.

Prominence can also be characterized at the level of the phonological phrase, with one or more words in the phrase carrying greater prominence than other words in the phrase. Phrasal prominence is typically identified with pragmatic focus in English. Prominence is assigned to words that introduce information that is new or important to the goal of the discourse or to words that bear contrastive focus (Bolinger 1986; Calhoun 2006; Selkirk 1996; Watson et al. 2008). A word that lacks prominence, on the other hand, must typically be *given* in the prior discourse context, i.e., anaphorically recoverable (Schwarzchild 1999). The relationship between prominence and focal status or givenness is especially strong for the rightmost prominent word (the *nuclear* prominence) in the phonological phrase, as shown by Calhoun (2006) in her large-corpus study of English. For pre-nuclear prominent words, prominence seems to depend more on phonological factors affecting rhythm, or on part-of-speech.

Phrasal prominence can also be modeled as structural prominence in phonological form by positing a layer of metrical structure above the word-level structure. To assign prominence to a word based on its focal status, or to avoid assigning prominence based on givenness, a speaker must deploy a prosodic phrase structure and metrical parse for the sentence that locates a word in the appropriately strong or weak position in metrical structure. Two sentences with the same words and syntactic structures can be pronounced with different prominence patterns, reflecting differences in the focal status or givenness of one or more words in the sentence.

Prominence as a phonological attribute is reflected in the phonetics in many ways. In English, phonetic effects of phrasal prominence are strongest in the lexically stressed syllable of the prominent word, which relative to non-prominent words exhibits hyper-articulation, increased duration and intensity, and increased spectral emphasis in the mid and high frequency regions (Beckman 1986; Beckman and Edwards 1994; Cole et al. 2007; Turk and White 1999; Cambier-Langeveld and Turk 1999; Kochanski et al. 2005; Sluijter and van Heuven 1996; Tamburini 2005). A prominent word may also be marked with a salient F0 movement expressing pitch accent (Pierrehumbert 1980; Ladd 1996), though corpus studies differ in finding evidence of F0 as a correlate of phrasal prominence (Kochanski et al. 2005; Calhoun 2006; Yoon 2007). We will refer to such properties collectively as marking “acoustic” prominence.

From the work just cited, we see that speakers assign structural prominence to words in a phonological phrase, taking into account the pragmatic and discourse properties of the words, and realize structural prominence as acoustic prominence

through increased duration, intensity, and prominence-lending F0 patterns. There is also evidence that listeners perceive the acoustic cues to prosody, and interpret the focal status or givenness of a word accordingly. Words with acoustic prominence are perceived by listeners as referring to new entities introduced in the discourse, or to entities with contrastive focus, while words with less acoustic prominence are perceived in association with prior discourse context or with universal givens (Arnold 2008; Dahan et al. 2002; Fowler and Housum 1987; Ito and Speer 2008).

The fact that (i) speakers encode focus and givenness in their phonetic productions, and (ii) listeners perceive focus and givenness in relation to acoustic prominence suggests a simple model of signal-driven prosody perception, in which listeners' perception of prosody and thus their interpretation of focus and givenness can be accurately predicted from cues present in the acoustic signal. But there are additional considerations that complicate the model. Beyond their role in signaling focus and givenness, the acoustic properties that signal prominence are also affected by factors related to the information status of a word. Specifically, words that are strongly predicted from the surrounding words or discourse context, and words that occur frequently in the language (high token frequency) have reduced acoustic prominence, as measured in duration, intensity, vowel formant dispersion and F0 (Aylett and Turk 2004; Bell et al. 2003; Fossler-Lussier and Morgan 1999; Gregory 2002; Ito, Speer and Beckman 2004; Munson 2007; Watson, Arnold and Tanenhaus 2008; Wright 2003). For example, Aylett and Turk (2004) show that in English, factors that encode the redundancy of a word predict 65% of the variance in raw syllable duration. In comparison, prosodic factors (as manually transcribed) predict up to 59% of the variance of the same duration measure.

The effect of word frequency on prominence is also evidenced in work that looks at effects of reduction using data from phonetic transcription. In English there is a close relationship between vowel quality and metrical structure: full vowels appear in metrically strong positions, while vowel reduction is characteristic of weak positions. Therefore, when a word is phonetically transcribed with a reduced vowel or no vowel in place of the vowel that appears in the corresponding dictionary form, the transcription reflects a low level of prominence for the affected syllable or word. Studies by Bybee (2001), Greenberg (1999), Greenberg and Fossler-Lussier (2000) and Bell et al. (2003) show that high-frequency words exhibit a greater incidence of consonant lenition and vowel reduction than low-frequency words. For example, Greenberg's (1999) study of spontaneous speech from the Switchboard corpus shows that compared to low-frequency words, high-frequency words have more pronunciation variants according to narrow phonetic transcription, reflecting a variety of reduction effects on consonants and vowels. Discussing the same corpus, Greenberg and Fossler-Lussier (2000) suggest that speakers intentionally modulate the precision of their articulation in relation to the entropy of an utterance: words associated with

low-entropy (highly predictable words) are produced with less precision, giving rise to more variation compared to high-entropy words.

Beside frequency, repetition and other factors related to the entropy associated with a word or phrase, there are at least two additional factors known to affect acoustic prominence. Wright (2003) and Munson (2007) find that words from sparse lexical neighborhoods are reduced relative to words from dense lexical neighborhoods, with evidence from measures of vowel dispersion. The final factor we mention is one that has long been known to affect acoustic prominence, and that is speaking rate. In their large-corpus study of factors influencing reduction, Fossler-Lussier and Morgan (1999) observe an increase in pronunciation variability, reflecting a greater incidence of segment reduction and therefore an overall lower occurrence of prominence as speaking rate increases.

Further complicating the relationship between the acoustic encoding of prosody and listeners' perception is the finding from Bard and Aylett's (1999) study of task-oriented dialogue, where words that are repeated have reduced acoustic prominence as expected, but are often still perceived as structurally prominent (*accented*, in the terminology of their study) by trained listeners performing prosody transcription.¹ This finding suggests that the perception of prominence is more complex than can be predicted by a simple signal-based model where acoustic cues are the primary influencing factor.

To summarize here, we have seen that the structural prominence of a word reflects its position in phonological metrical structure, and that factors related to focus and givenness also play a role in determining the assignment of prominence at the phrase level, especially for nuclear prominence. Prominence is encoded in the acoustic signal in duration, intensity, spectral emphasis, and F0 and in measures related to vowel dispersion. Furthermore, perception studies show that listeners perceive acoustic prominence in relation to focus and givenness. At the same time, structural prominence, whether reflecting focus status or phrase-level metrical structure, is not the only determinant of acoustic prominence. Non-structural factors, in particular, the predictability of a word, the density of its lexical neighborhood and speech rate also condition variation in many of the same acoustic parameters that correlate with phonological prominence as an expression of focus and givenness. A further complication is that the structural and non-structural factors influencing acoustic prominence may not be wholly independent of one another. For instance, the fact that high-frequency words such as function words typically have low acoustic prominence may be related to their common occurrence in positions of low structural prominence. High-frequency words are more predictable and therefore less likely to introduce important new information to the discourse. Consequently, high-frequency words are less likely to carry pragmatic focus and thus, less likely to be assigned structural prominence (i.e., less likely to bear a pitch accent). Conventionalization of their acoustic form from positions of low structural prominence may be the basis for their overall tendency to have low acoustic prominence.

Clearly, the patterns of acoustic variation in duration, intensity, F0 and spectral measures are very complex, which leads us to our research question: *How do listeners perceive prominence in everyday speech?* Our focus is on prominence as it naturally occurs in spontaneous speech produced in genuine communicative contexts. We believe that the interactive communications of everyday speech are a good place to look for a broad range of expressions and focus conditions. Ordinary, spontaneous speech is also rich in the kinds of reduction associated with speech rate and entropy discussed above. In short, spontaneous speech presents all the factors that challenge our understanding of prominence production and perception. Our approach is to study listeners' perception of prominence broadly construed. We do not set out to isolate structural from non-structural prominence through control of experimental materials, but consider all instances of prominence as judged by listeners, and investigate the basis of perceived prominence in structural and non-structural factors collectively. The present study, in particular, asks about the influence of non-structural factors on listeners' perception of prominence in conversational speech.

The study presented here is based on speech data from the Buckeye corpus of conversational speech collected through face-to-face interviews (Pitt et al. 2007). Excerpts from this corpus were transcribed for the occurrence of prominence by ordinary, untrained listeners. Section 2 introduces the corpus, the transcription experiment, and results from our transcriber reliability study. This study investigates patterns of prominence perception in relation to the acoustic correlates of prominence, and in relation to the word's information status, and also examines the relationship between acoustic correlates of prominence and a word's information status. The goal of the study is to understand the basis of prominence perception in acoustic cues and in the information status (i.e., predictability) of a word. We also want to know if acoustic cues and information status are distinct in their influence on prominence perception, or whether they converge on a common prominence judgment for a word. The three facets of this study – perception, production, and information status – are laid out schematically in Figure 1 in relation to the sections of the paper that present the experimental findings. Section 3 contains correlation and regression analyses of the relationship between perceived prominence and acoustic measures of prominence, and between perceived prominence and factors related to a word's information status. Section 4 presents correlation and regression analyses relating acoustic prominence and factors of information status, and Section 5 extends the statistical analysis with hierarchical and non-linear regression models. From the experimental findings in sections 3, 4 and 5 we conclude that information status, and word frequency in particular, influences prominence perception at least partly independently of the acoustic properties of a word. Section 6 introduces a processing model that accounts for the contribution of information status to prominence perception. In this model, prominence perception is both signal-driven, based on the speaker's phonetic implementation of prominence, and expectation-driven, based on the listener's prior experience.

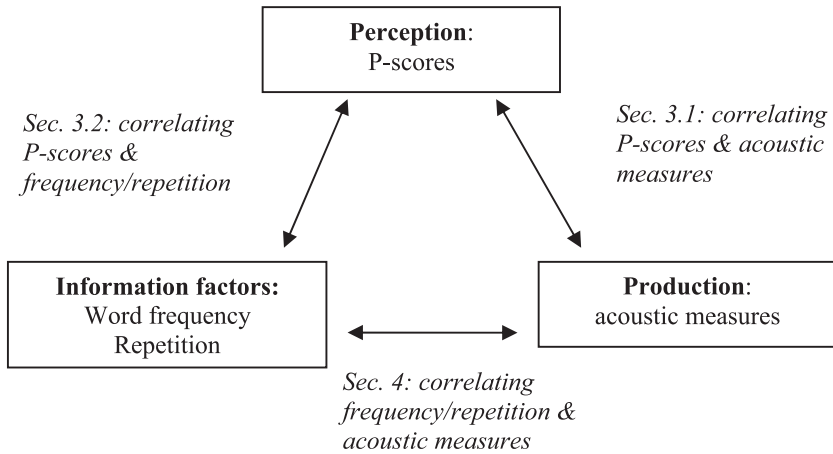


Figure 1. The design of the present study, relating the listener's perception of prominence, the speaker's production of prominence, and factors related to the information status of a word.

2. Experiment in naïve prosody transcription

In order to explore the correlates of prominence in acoustic or other properties, it is necessary to construct a speech database that is annotated for the location of prominence. Most prior work on prosody relies on one of two methods for collecting prosody data. The laboratory method involves engaging subjects in controlled speech tasks, often using read speech, where the tasks are carefully designed to elicit the range of prosodic events (e.g., prominent words, prosodic phrase boundaries) under investigation. An alternative method gathers speech samples from a corpus where speakers were not explicitly instructed nor materials explicitly designed for the purpose of eliciting specific prosodic events. These speech materials are then subject to prosodic annotation, most often conducted by one or a small number of highly trained experts, typically including the experimenter. Both of these methods have yielded significant insight into speech prosody, but leave open certain questions such as whether the prosodic patterns in controlled laboratory speech are replicated in ordinary speech, or whether the prosodic properties transcribed by a single expert listener are the same properties that any ordinary, “naïve” listener would perceive and interpret with respect to the pragmatics and discourse structure of the utterance.

2.1. Materials

To obtain a speech prosody database representative of natural, spontaneous speech we draw on the Buckeye corpus of conversational speech with speakers from Co-

lumbus, Ohio (Pitt et al. 2007). Excerpts were extracted from the interviews with 37 speakers. One speaker was used for demonstration purposes. Two short excerpts of 11–22 seconds long from each of the remaining speakers were extracted, for a total of 72 short excerpts. A longer excerpt of 31–58 seconds in duration was extracted from 18 of the same speakers. Transcription experiments with the long excerpts were conducted in order to look at the effect of repetition within the discourse segment. Word transcriptions, in regular orthography, were taken for each excerpt from the transcriptions published with the corpus. Transcriptions were modified to remove all punctuation and capitalization.

2.2. Method

The speech excerpts were subject to a coarse prosody transcription using untrained transcribers who were naïve to the questions and methods of prosody research. Between 15–22 transcribers transcribed each speech excerpt, and the results were pooled over transcribers to obtain a population-wise, probabilistic measure of the prosodic status of each word. This method is adapted from similar methods used by Buhmann et al. (2002), Swerts (1997), and Streefkerk et al. (1997, 1998).


A total of 97 listeners recruited from undergraduates at the University of Illinois participated in the prosody transcription study. The data from 6 listeners were excluded due to failure to follow the transcription guidelines or because they were found not to be monolingual. Listeners included in the analyses had no prior training in prosody transcription or prosodic phonology and were monolingual, native speakers of American English. The majority were residents of Illinois from childhood, from areas associated with the Northern Cities or Midlands varieties of American English. Subjects were seated at computers and given minimal instruction. They were told that they would hear excerpts from recorded interviews with speakers of American English, and they would mark the transcript for each word they heard as prominent. The experimenter defined prominence by reading the short script in (1), but no example sound file was played to demonstrate prominence. Subjects were also told to expect variation among speakers in the frequency and expression of prominence, and that they should not be concerned with getting the “right” transcription, as the experimenter was interested in how listeners might differ in their perception of prominence.

(1) Instruction script read to subjects:

“In normal speech, speakers pronounce some word or words in a sentence with more prominence than others. The prominent words are in a sense highlighted for the listener, and stand out from other non-prominent words. In some of the excerpts you will hear, you will be asked to mark all prominent words by underlining them.”

The experiment was run in five sessions, with a total of 74 subjects randomly assigned to one of four subsets of short excerpts, with between 15–22 subjects per

subset, and an additional 23 subjects assigned to the long excerpts. All subjects within a subset were presented with the same audio files in randomized order. The ordering of excerpts on the printed transcripts followed the order of audio presentation. Subjects listened to the speech excerpts through headphones and were asked to mark the printed transcript for the location of prominent words. The experiment also included a parallel task where the same subjects transcribed a different subset of speech excerpts for the location of perceived prosodic phrase breaks. Subjects did not view any graphical display of the speech signal, so transcriptions were based only on auditory impression. Subjects worked through the transcription task at their own pace, opening sound files in a fixed sequence by clicking on icons appearing on the screen. The transcriptions were done in real time and subjects could not stop or re-start the recording. They listened to each excerpt twice through in succession, and were allowed to mark changes to their transcription on the second play as shown schematically in (2). Prominent words were marked with an underline beneath the entire word (2a). A mark could be retracted by striking through the word (2b), and a retracted mark could be recalled by circling the word (2c). No further changes could be recorded. Subjects were instructed not to attempt erasing, in order to avoid slowing down and losing track of the flow of speech.

- (2) a. word word word
 b. word ~~word~~ word
 c. word  word

2.3. Data coding and assessing reliability

The transcriptions were pooled over all listeners to obtain two population-wise prosody scores for each word. The P-score is a number between 0 and 1 that represents the proportion of transcribers who perceived that word as prominent, and the B-score is a similar encoding for the perception of a boundary following the word. These scores provide a probabilistic coding of prosody for each word, with higher scores indicating greater agreement among transcribers, presumably reflecting less ambiguity in the prosodic organization of the utterance, and/or the presence of stronger or more salient cues. Figure 2 shows an example of a partial excerpt, plotting the P-score and B-score for each word. This example illustrates a typical finding, namely, that there are many words that transcribers agree are *not* prominent (or *not* at a boundary), but there are few if any words where transcribers reach the same rate of agreement on the positive assignment of prominence (or boundary).

The reliability of this probabilistic coding of prosody was assessed through a study of inter-transcriber agreement for the short excerpt transcriptions. Similar results are expected for the long excerpts. Given the large number of transcribers marking each set of excerpts, we used Fleiss' multi-rater agreement statistic (Fleiss 1971), rather than the more frequently used Cohen's statistic, which is calculated

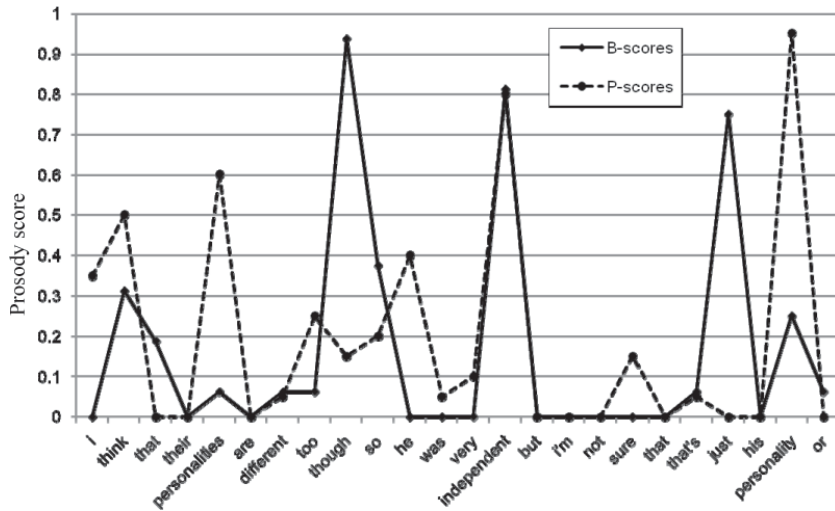


Figure 2. Graph of probabilistic prominence (P) scores and boundary (B) scores for each word in a sample utterance from the test corpus. Prosody scores are based on pooled transcriptions of 20 transcribers.

Table 1. Results from the multi-transcriber reliability study for the transcriber groups assigned to each set of short excerpts, with Fleiss' kappa coefficients and their normalized z-scores. At $\alpha = 0.01$, significance is reached at $z = 2.32$. All z-scores are highly significant.

Excerpt set		1	2	3	4
prominence	Kappa	0.373	0.421	0.394	0.407
	z	19.43	20.48	18.15	18.31
boundary	Kappa	0.612	0.544	0.621	0.575
	z	27.62	21.87	25.05	26.22

over pairs of raters. Like Cohen's statistic, Fleiss' statistic measures the actual agreement in relation to the expected agreement based on the number of raters and the number of response options, but factors-in response variation over the entire group of raters in the calculation of expected agreement. Fleiss' kappa coefficients and their normalized z-scores are shown in Table 1. Agreement rates are higher for boundary perception than for prominence, but the z-scores are in every case highly significant ($\alpha = 0.01$), indicating that transcriber agreement in both boundary and prominence perception is reliably beyond chance expectation. Further results from the reliability study are reported in Mo, Cole and Lee (2008).

The results from Fleiss' statistic, and similar findings from Cohen's kappa statistic over all pairs of transcribers (reported in Mo, Cole and Lee 2008) indicate variability in listeners' perception of prominence and boundary. The probabilistic

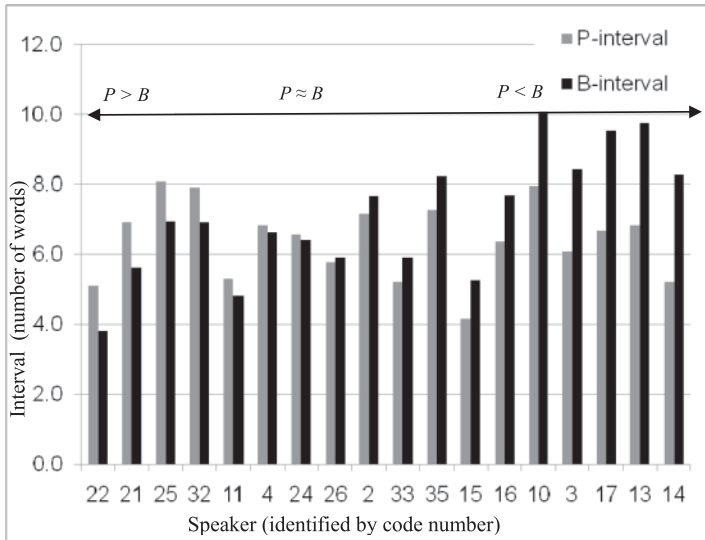


Figure 3. Plot of the mean interval between perceived prominences (*P*-interval) and boundaries (*B*-interval) for 18 speakers, shown individually. Intervals are measured in number of words. Mean intervals are calculated based on all transcribers who transcribed a given speaker (15–22), and all transcribed utterances from that speaker. Speakers differ in the relative length of *P*-intervals and *B*-intervals, as described in text.

prosody scores also reveal variability across speakers. Looking at transcribers' responses by speaker, we observe that the same transcribers respond differently to different speakers. Figure 3 shows the pattern of *P*-scores and *B*-scores for the same group of transcribers for each of 18 speakers, in a plot of the mean interval between prominences or boundaries over all the transcribed utterances for a given speaker. The mean interval measure is a measure of the frequency of prominent words and boundaries for each speaker, as perceived by the group of 15–22 transcribers who transcribed each speaker. The speakers displayed in the center of the chart are those for whom transcribers perceive roughly equal intervals between prominences and boundaries (labeled on the chart as $P \approx B$), indicating that on average there is one prominence within each perceived prosodic phrase. We interpret this as a pattern in which transcribers are mostly responding to the nuclear accented word within the phrase (typically the rightmost accented word). Speakers towards the left side of the chart are those for whom transcribers perceive on average a shorter interval between boundaries than between prominent words ($P > B$). This means that there are some intervals for which transcribers perceive no prominent words. Speakers towards the right side of the chart are perceived on average as producing longer intervals between boundaries than between prominences ($P < B$), indicating that transcribers perceive multiple prominences in at least some prosodic phrases for these speakers. This variation in the patterns of utterance-

level prosody perception by speaker suggests that speakers may differ from one another either in their use of prosody (i.e., in the phonological prosodic structures assigned to an utterance), or in the salience of their phonetic encoding of prosody, or in both, and that listeners are sensitive to these individual differences in prosody.

3. Correlates of perceived prominence

The reliability study shows that listeners agree at above chance levels on their perception of the prominence status of words in conversational speech, and in this section we explore the bases of prominence perception, testing two hypotheses. First, we examine the acoustic properties of words in relation to their perceived prominence, to test the hypothesis that prominence perception is *signal-driven*, based on the acoustic encoding of phonological prosodic features. The second hypothesis we test is termed *expectation-driven perception*, which is based on the finding from prior studies that the information status of a word correlates with acoustic prominence. If a listener recognizes a word as low-frequency or repeated from an earlier mention in the discourse, then based on their prior experience with such words, the listener may simply expect that such a word will not be prominent, independent of its actual acoustic form, and judge it as such in the transcription exercise. Under this scenario, the listener may judge prominence based on information status alone, rather than judging the acoustic form directly. In this sense, the listener's judgment of word prominence is driven by their expectation based on prior experience of the word. There is a third hypothesis that is tested in Section 4, and that is that the acoustic signal and the information status of a word mutually predict the listeners' judgment of prominence. This would be the situation if acoustic prominence is correlated not only with perceived prominence, but equally with a word's information status, e.g., if high-frequency and repeated words have the same characteristics of low acoustic prominence.

3.1. *Acoustic differences related to perceived prominence*

Acoustic measures of intensity, duration, spectral emphasis, and less consistently F0, have been shown in prior work to correlate with phrasal prominence. We have examined the same parameters from all stressed vowels of each word from the short excerpts, using correlation analyses with measurements that were z-normalized within each vowel phone category. Table 2 shows each vowel and its frequency in our database of short excerpts.

Results of correlation analyses for duration (ms), RMS intensity (dB), and RMS bandpass filtered intensity (dB) as predictors of perceived prominence for each stressed vowel are reported in Mo (2008) and summarized here.² F0 measures yield no or very weak relation to P-scores under ANOVA and correlation analyses, and are not discussed further here.³ The bandpass filtered intensity measures were

Table 2. The stressed vowels used for acoustic analysis and their number of instances in the short excerpts, including all 18 speakers.

vowel	ɑ	æ	ʌ	ɔ	əʊ	aɪ	ɛ	ɜː	eɪ	ɪ	i	oo	ʊ	u
N	81	129	211	58	28	140	187	66	114	209	156	103	41	94

Table 3. Pearson's correlation analyses of P-scores and acoustic measures taken from the stressed vowel of each word. Correlation coefficients (Pearson's r) are shown here for correlations run separately for each vowel phone. Significant correlations (1-tailed) are marked with * ($p < .05$) and ** ($p < .01$).

vowel	ɑ	æ	ʌ	ɔ	əʊ	aɪ	ɛ	ɜː	eɪ	ɪ	i	oo	ʊ	u
Duration	.033	.301**	.198**	.224*	.491*	.419**	.237**	.160	.302**	.244**	.266**	-.128	-.042	.141
Overall	.304**	.114	.147*	-.043	.151	.209**	.220**	.283**	.137	.228**	.139*	.123	.308*	.005
Intensity	.174	.098	.078	-.096	.076	.184*	.137*	.237*	.116	.210**	.138*	.140	.268*	.005
0–.5 kHz														
Intensity	.343**	.163*	.271**	.041	.209	.238**	.282**	.349**	.093	.262**	.105	.120	.328*	.024
.5–2 kHz														
Intensity	.175	.263**	.150*	.001	.098	.152*	.264**	.035	.184*	.201**	.132	-.018	.177	-.056
2–4 kHz														

taken in the frequency bands 0–.5, .5–2, 2–4 kHz to assess spectral emphasis, shown by Sluijter and van Heuven (1996) to be a reliable correlate of word stress in Dutch, and by Heldner (2003) and Tamburini (2005) also to be characteristic of phrasal stress prominence in Swedish and English, respectively.

Correlation analyses were conducted to explore the relationship between perceived prominence and acoustic measures. Pearson's bivariate correlations between P-scores and the acoustic measures of duration and overall intensity were calculated over all vowels. Both measures show a significant positive correlation [Duration: $r = .204$, $p < .001$; Overall intensity: $r = .180$, $p < .001$]. More listeners perceive a word as prominent when its stressed vowels are longer in duration and louder, though the correlation with duration is stronger. Table 3 shows correlation results for individual vowels. Again, we observe many correlations between P-scores and duration or intensity measures, but over the set of 14 vowels, more vowels show a significant correlation with duration than with the intensity measures, and the correlation strength is greater for duration than for any intensity measure. Among vowels, the maximum r for duration is 0.491, while the next highest correlation is with Intensity .5–2 kHz, where $r = 0.349$. The breakdown by vowel also reveals that for the non-low back, rounded vowels /oo, u/ no correlation holds between P-scores and measures of acoustic prominence. We do not pursue this finding further, but consider the possibility that some vowels are inherently more prominent than others.⁴

These findings reveal a pattern of acoustic correlates of perceived prominence in our study that resemble the patterns reported in prior studies. In our study, untrained listeners perceive a word as not prominent when it has weak acoustic prominence, and are most consistent in judging a word as prominent when it has enhanced acoustic prominence. This pattern is strongest for the acoustic correlate of duration.

3.2. *Information status correlates of perceived prominence*

Correlation analyses were conducted to test the relationship between the information status of a word and listeners' perception of prominence. Two measures of information status were evaluated. First, we estimated the token frequency of each word in our corpus with the log frequency of the same word in the Switchboard corpus of spontaneous, conversational speech.⁵ The Switchboard corpus is much larger than Buckeye, comprising over 240 hours of recorded telephone conversations from over 500 speakers of American English (Godfrey, Holliman and McDaniel 1992), and therefore provides a better basis for calculating token frequency. The second measure of information status was the repetition index encoding the number of times a word had been repeated in the preceding portion of its discourse segment, i.e., in the 31–58 s. "long" excerpts used for this part of the analysis. The first mention of a word has the repetition index of 1, second mention has the index 2, and so on. Because repeated words are less common in shorter excerpts, we expanded the database for this part of the study to include the long

Table 4. *Pearson's correlation and linear regression analyses of P-scores (pooled over transcribers) and log-frequency (calculated from the Switchboard corpus), for words in four data sets: all short excerpts, short excerpts minus frequently reduced words or minus function words, and long excerpts. Significant correlations (r) at $p < .001$ (1-tail) are marked with **.*

Data set	N	Pearson's r	r^2
Short excerpts, all words	2024	-.505**	.255
Short excerpts minus frequently reduced words	1217	-.432**	.187
Short excerpts minus function words	778	-.302**	.091
Long excerpts	1725	-.432**	.187

excerpts from 9 speakers. Correlation analyses are reported for the short and long excerpts separately in what follows. Furthermore, since function words tend to be repeated in a discourse segment more often than content words, reflecting the overall high token frequency of function words, we conducted additional correlation analyses on the set of repeated words after removing all function words from the set. For similar reasons, we also analyzed the data set after removing frequently reduced words (including many pronouns, determiners, auxiliary verbs, prepositions, and sentence joiners), from a list of about 80 items identified by Huddleston and Pullum (2002). We reason that frequently reduced words may be judged for prominence differently than words that are not frequently reduced, especially if the reduced form represents a distinct lexical item than its less common unreduced counterpart.

Looking first at log-frequency, Table 4 shows a significant negative correlation between log-frequency and P-score. Listeners are more likely to perceive a word as prominent (yielding a higher P-score for the word) if it is a low-frequency word. Linear regression analysis shows higher r^2 values for the dataset of short excerpts that includes all words, indicating that frequency is a stronger predictor of perceived prominence when function words and frequently reduced words are included in the analysis. This is not surprising, given that function words have high frequency and are often reduced, and reduction is a characteristic of non-prominent words. When such words are removed, the r^2 values drop, indicating that for most content words factors other than word frequency play a larger role in prominence perception than they do for function words. But even without function words and reduced words, the correlation between word frequency and P-scores remains highly significant. The long excerpts show a similar pattern of negative correlation between word frequency and P-scores, but with r^2 values that are intermediate between those of the various datasets for the short excerpts. This finding most likely reflects the fact that the long excerpts were chosen to provide more data on repetition effects, with the criterion that they included a large number of repeated words, and especially repeated content words. Thus, it is reasonable to think that the influence of function words and reduced words on prominence perception may have been less for the long excerpts.

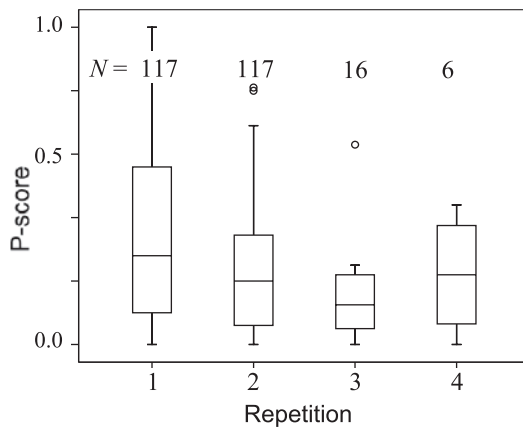


Figure 4. Boxplots of P-scores for words in short excerpts, grouped by repetition index. These plots include only words that occur with at least two instances (repetition indices 1 and 2) in the same discourse segment. The number of words in each group is indicated above each box. Total $N = 256$ words.

Turning now to the relationship between word repetition and perceived prominence, Figure 4 shows the distribution of P-scores for words in the short excerpts that are repeated at least once in the excerpted part of the discourse segment. There is a trend towards decreasing P-scores from the first to third mention of a word, but the P-scores increase for the fourth mention. We observe a similar trend of increasing P-scores after the second or third mention of a word in every dataset we have examined, and the effect seems to reflect an increased prominence on a word that is reintroduced in the discourse, sometimes with contrastive focus. It bears noting that the number of words in each repetition group decreases sharply after the second mention, indicating that while repetition is common in these materials, there are still few instances where a word is repeated more than two times within a single discourse segment.

Table 5 shows the results of correlation and regression analyses for the factors P-score and Repetition index. Again, we look separately at short excerpts, with and without function words, and long excerpts. For the short excerpts, there is a significant negative correlation between P-score and Repetition index in the expected direction: repeated words tend to be perceived as prominent less consistently than first mention words. The correlation is stronger over the first through third mention, and decreases somewhat when fourth and subsequent mention words are included in the model. This reflects the upward trend of P-scores for fourth (and subsequent) mention words, as seen in Figure 4. The same trend is evident in the P-scores for repeated words in the long excerpts. The correlation between P-scores and Repetition index for the long excerpts is not significant when all repetitions are included in the model, but when we restrict the model to just first and second mention words, there is a significant negative correlation.

Table 5. *Pearson's correlation and linear regression analyses of P-scores (pooled over transcribers) and repetition indices for words in short excerpts (with and without frequently reduced words and function words) and long excerpts. Significant correlations (r) at $p < .001$ (1-tail) are marked with **.*

Data set	Repetition coding	N	Pearson's r	r^2
Short excerpts, all words	1 st –6th repetition	891	-.113**	.013
	1st vs. 2nd vs. 3rd+	891	-.128**	.016
Short excerpts minus function words	1st–4th repetition	164	-.242**	.059
Long excerpts, all words	1 st –5th+ repetition	481	-.061	.002
Long excerpts, all words	1 st vs. 2 nd repetition only	299	-.139**	.017

The correlation analyses presented in this subsection show that listeners' perception of prominence is related to word frequency, and to a lesser extent, the word's status as repeated in the discourse segment. Word frequency is a better predictor of perceived prominence than the repetition index, accounting for 26% and 19% of the variance in P-scores in the short and long excerpts, respectively.⁶ Function words and other frequently reduced words contribute strongly to this effect, but the correlation is still fairly strong for long excerpts, which were chosen to maximize repeated content words.

3.3. *Interim summary*

The correlation analyses in Section 3.1 show that the perception of prominence in conversational speech by ordinary listeners is correlated with acoustic measures of prominence, consistent with the hypothesis of *signal-driven prominence perception* and the findings reported in prior work. Section 3.2 shows that prominence perception is also correlated with word frequency and to a lesser degree repetition. In Section 1 we reviewed prior work showing a relationship between acoustic prominence and measures of information status, such as word frequency. Frequent words, and words that are strongly predicted by context, have lower acoustic prominence compared to low-frequency and less predictable words. This link between information status and acoustic prominence raises a question for our study. Does acoustic prominence (in duration and intensity) correlate with word frequency and repetition in the Buckeye data, and if so, does it explain the correlation we observe between word frequency and perceived prominence? In other words, are the effects of word frequency and repetition on perceived prominence in this study modulated through acoustic information? If so, then we could say that listeners judge prominence on the basis of the acoustic signal, perceiving words as prominent when they have increased duration and intensity. Any factor that influences the acoustic correlates of prominence would be expected to correlate with perceived prominence, but only indirectly. This is the scenario of signal-driven prominence perception. On the other hand, if word frequency or repetition were found not to influence duration and intensity, contrary to expectations based on

findings from prior studies, that would imply that the correlation between perceived prominence and word frequency or perceived prominence and repetition is not mediated by acoustic properties, and instead, that listeners are responding directly to the information status of a word in judging its prominence.

To further probe the role of a word's information status in prominence perception, we turn next to an analysis of the correlation between acoustic measures and the measures of information status – word frequency and repetition. The acoustic correlates of frequency and repetition will be compared to the acoustic correlates of perceived prominence, to further test the hypothesis of signal-based prominence perception.

4. Acoustic correlates of word (log-) frequency and repetition

Correlation and regression analyses were conducted to test the relationship between acoustic measures of prominence and the information status features of log-frequency and repetition. An interesting complementarity appears between the acoustic correlates of frequency and repetition. Duration is the only significant acoustic correlate of repetition ($r = -0.15$, $p < .05$), whereas intensity and spectral emphasis are correlates of word frequency, but duration is not (Overall intensity: $r = -0.15$, $p = .041$; Intensity 0–.5 kHz: $r = -0.18$, $p < .05$; Intensity .5–2 kHz: $r = -0.17$, $p < .05$; Intensity 2–4 kHz: $r = -0.67$, $p < .01$). Linear regression statistics are given in Table 6, and with one exception, the low r^2 values indicate that log-frequency and repetition are not very strong predictors of variation in these acoustic prominence measures. The exception is with log-frequency, which is a fairly good predictor of intensity in the 2–4 kHz frequency band ($r^2 = .449$).

If the correlations reported in section 3.2 between perceived prominence and measures of information status (frequency and repetition) are modulated through acoustic information, then we would expect to find similarities between, for example acoustic correlates of prominent and low-frequency words, or between acoustic correlates of prominent and first mention (or not-repeated) words. We

Table 6. *Linear regression statistics (r^2) for word log-frequency and repetition as predictors of acoustic measures of duration, intensity and spectral emphasis (bandpass-filtered intensity). Data from short excerpts, including only words that have repeated mention. Acoustic measures from all stressed vowels, pooled. Asterisks mark cells where the correlation between the two factors is significant at $p < .05$ (1-tail).*

	Log-frequency	Repetition
Duration	0.001	0.024*
Overall intensity	0.024*	0.001
Intensity 0–.5 kHz	0.034*	0.002
Intensity .5–2 kHz	0.028*	0.003
Intensity 2–4 kHz	0.449*	0.002

Table 7. *Comparison of significant acoustic correlates of word frequency (log-frequency) and perceived prominence (P-scores) from analyses of short excerpts. Mean r^2 values for all vowels are shown in parentheses, summarizing data from Tables 3 and 6.*

		Frequency	Perceived prominence
Significant acoustic correlates	duration	No	Yes ($r^2 = .089$)
	intensity	Yes ($r^2 = .024$)	Yes ($r^2 = .053$)
	spectral emphasis	Yes ($r^2 = .449$)	Yes ($r^2 = .07$)
		(high frequency)	(mid frequency)

focus our comparison on acoustic correlates of prominence vs. frequency, leaving aside repetition for the moment, since we saw in section 3.2 that frequency is more strongly correlated with perceived prominence than is repetition.

From Table 3 above we observe that among the acoustic measures, duration and intensity in the mid-range frequency band (.5–2 kHz) are the most reliably correlated with P-scores: 9 out of 14 vowels show duration as a correlate of P-scores with r values ranging from 0.198–0.419; 8 vowels show mid-range intensity as a correlate of P-scores with r values ranging from 0.271–.349.⁷ On the other hand, we see from Table 6 that overall intensity and especially spectral emphasis (with increased energy in the high frequency range) are the primary correlates of word frequency. With the exception of spectral emphasis as a correlate of word frequency, these statistically significant correlations are weak, accounting for less than 10% of the variance in the dependent variable (P-scores or acoustic measures). The weak correlations likely reflect the fact that these data are drawn from a corpus of spontaneous speech. We have not controlled for (or statistically modeled) acoustic variability due to individual speakers or due to contextual factors known to influence these acoustic measures. Nonetheless, comparing these findings we may say that words perceived as prominent tend to have longer duration and a weak tendency for spectral emphasis in the mid frequency range, while words that are low-frequency do not show these effects, but show a much stronger tendency for increased high-frequency spectral emphasis instead. Table 7 summarizes the comparison. Perceived prominence and word frequency differ somewhat in their acoustic characterization, which suggests that the correlation we have observed between perceived prominence and word frequency is not completely modulated through acoustics. The effect of word frequency on prominence perception appears to be at least partly independent of acoustic prominence.

5. Statistical models of P-score variance

To further explore the role of acoustic measures, word frequency and repetition as factors that influence listeners' perception of prominence, we conducted additional statistical analyses with regression using Hierarchical Linear Models (HLM).

Table 8. *Results of hierarchical linear regression models with acoustic measures, word log-frequency and repetition index as predictors of P-score variance. Models I and II differ in the order in which the factors were applied.*

			R ²	R ² change	Sig. of R ² change
With acoustic measures	Model I	Log_freq & rep	.187	.187	<.001
		Dur	.245	.058	<.001
		Intensities	.269	.024	<.001
	Model II	Dur	.064	.064	<.001
		Intensities	.093	.030	<.001
		Log_freq & rep	.269	.175	<.001

HLM allows us to test the individual contribution of each factor, or sets of factors, in accounting for the overall variance in P-scores. Individual factors are entered in the model in a step-wise fashion (in separate levels of the model), as predictors of P-score variance. Analyses with HLM were carried out with the full dataset from the short excerpts.

The first two models, summarized in Table 8, coded three levels of predictor variables: the normalized duration measure of a word was entered by itself as one level, the log frequency and repetition index of a word were combined in a second level, and the intensity measures (overall intensity, and three sub-band intensity measures) were combined in a third level. In Model I the information status measures (log frequency and repetition index) were entered in the first step, followed by duration and intensity measures in the second and third steps, respectively. This model simulates a perceptual process by which the information status of a word impacts the perception of prominence before the acoustic prominence is considered. Model II starts with the acoustic measures in the first two steps (duration followed by intensity), and then enters the information status measures in the final step, simulating a perceptual process in which acoustic prominence is considered prior to word frequency and repetition. As shown in Table 8 factors at all levels were significant predictors of P-score variance. The R^2 change values show that the information status measures taken together were the strongest predictors of P-score variance, accounting for 19% of the variance in Model I (when applied first) and 18% of the variance in Model II (when applied last). The duration measure was the second largest predictor variable, accounting for 6% of the overall P-score variance in both Models I and II. Intensity measures combined were the least predictive factors in these models, accounting for a mere 2% (Model I) or 3% (Model II) of the total variance. These models achieved comparable results, both predicting a total of 27% of the variance in P-score values.

Additional tests using HLM were conducted using Principal Component Analysis (PCA) to further reduce the number of predictor variables, eliminating redundancy in the sets of factors combined into a single level in Models I and II. PCA selects all components with eigenvalues greater than one, which in each case

Table 9. *Results of hierarchical linear regression models using Principal Component Analysis with acoustic intensity (Models III, IV, VII, VIII) and with word log-frequency and repetition index (Models IV, V, VII, VIII) as predictors of P-score variance. Pairs of models with the same predictor variables differ in the order in which the predictor variables are applied.*

			R ²	R ² change	Sig. of R ² change
With PCA of intensity	Model III	Log_freq & rep	.187	.187	<.001
		Dur	.245	.058	<.001
		PC (intensity)	.260	.015	<.001
	Model IV	Dur	.064	.064	<.001
		PC (intensity)	.080	.018	<.001
		Log_freq & rep	.260	.179	<.001
With PCA of word info	Model V	PC (info)	.126	.126	<.001
		Dur	.185	.059	<.001
		intensities	.214	.029	<.001
	Model VI	Dur	.064	.064	<.001
		intensities	.093	.030	<.001
		PC (info)	.214	.121	<.001
With PCA of intensity and word info	Model VII	PC (info)	.126	.126	<.001
		Dur	.185	.059	<.001
		PC (intensity)	.202	.017	<.001
	Model VIII	Dur	.064	.064	<.001
		PC (intensity)	.081	.018	<.001
		PC (info)	.202	.121	<.001

presented here resulted in a single component. Models III and IV were tested using PCA over the set of intensity features. The resulting intensity component was entered by itself in a separate level of the HLM, with the combined frequency and repetition measures as a second level and duration as a third level. As before, Model III first applies frequency and repetition, followed by the acoustic measures of duration and the principal component measure of intensity in steps two and three. Model IV uses the same levels, applied in the sequence of Model II (acoustic predictors precede information status predictors). Table 9 shows that for Models III and IV, as for Models I and II, the combined factors of frequency and repetition are the strongest predictors of P-scores, with duration and PC-intensity following in order. The overall success of these models using the principal component of intensity measures is slightly less than for Models I and II, accounting for 26% of the variance of P-scores.

Models V and VI were constructed using PCA to reduce the two measures of a word's information status to a single factor. As seen in Table 9, these models are less successful than Models I–IV, accounting for only 21% of P-score variance, due to the weaker contribution from the principal component factor based on log-frequency and repetition. Models VII and VIII fare even worse, using principal components of word information status *and* intensity, and accounting for a total of 20% of P-score variance.

Table 10. *Comparison of results from linear, quadratic and cubic regression models predicting P-scores (from short excerpts) from three sets of factors: word log-frequency and repetition; duration, and the principal component of intensity measures.*

		R ²	Sig.
Word_freq & rep	Linear	.187	<.001
	Quadratic	.187	<.001
	Cubic	.195	<.001
Dur	Linear	.061	<.001
	Quadratic	.066	<.001
	Cubic	.070	<.001
PC (intensity)	Linear	.021	<.001
	Quadratic	.021	<.001
	Cubic	.022	<.001

These HLM results indicate that although we can reduce the complexity of the analysis and eliminate potential redundancy in the predictor variables by using PCA, the resulting models are not as successful in predicting P-scores as are the models that code each factor individually. This finding suggests that each of the individual factors considered here contributes in some way to listeners' perception of prominence. A further observation from these models is that the contribution from the duration measure was relatively stable across models. This indicates that duration plays a moderate but consistent role in signaling prominence despite differences in the role of other factors due to method of analysis.

The HLM tests use linear regression, which are appropriate models under the assumption that the predictor variables are linearly related to P-scores. In a recent study examining the relationship between word frequency and acoustic duration, Kuperman et al. (2008) demonstrate that non-linear functions provide a better model of the relationship between duration and word frequency in spontaneous speech data from Dutch, German, Italian and American English. Indeed, they use a subset of the Buckeye corpus for their analysis of American English, which is the same corpus used in the present study. To investigate the possibility of a non-linear relationship between P-scores and information status measures, we conducted a series of non-linear regression analyses. Specifically, we compared results between regression models using linear, quadratic and cubic functions. The results of this comparison are given in Table 10 and show that the cubic model uniformly provides the best prediction of P-scores with each of the three sets of predictor variables. This finding is consistent with the results of Kuperman et al. (2008), where words with low and high duration tend to have lower log-frequency than words with less extreme duration values. Yet, the cubic models also have greater complexity (a greater degree of freedom). Comparing the F-ratio of the cubic and linear models, the increase of the sum of squares divided by the increase in the degree of freedom reveals no significant gain for the cubic over the linear model. We conclude that

although the non-linear cubic model provides a somewhat better fit for the relationship between P-scores measures of acoustic prominence and the information status of a word, the linear model offers a comparable and simpler model. We leave further investigation of non-linear models for future work.

To summarize, the findings from the HLM and non-linear regression models provide strong evidence for non-structural factors related to a word's information status as factors that influence listeners' perception of prominence in spontaneous speech. Under all the models tested the information status factors are stronger predictors of perceived prominence, by a factor of two or more, compared to the acoustic factors of duration, overall intensity and spectral emphasis. All of the factors are significant predictors of perceived prominence, and the HLM models clearly show that no factor is redundant: each factor is significant in each model, regardless of the order of application. These findings demonstrate that prominence perception is both signal-driven (influenced by acoustic factors) and expectation-driven (influenced by word frequency and repetition).

6. A processing model of factors influencing prominence perception

What is the mechanism by which word frequency, and possibly other factors related to the information status of a word, can influence a listener's judgment of the prominence of a word in an utterance? We propose that this effect can be modeled with reference to lexical processing. The basic idea is that the listener's judgment of prominence may directly reflect the speed or ease of lexical access.

As discussed in Section 1, word predictability is inversely related to acoustic prominence (Bell et al. 2003; Gregory 2002; Watson, Arnold and Tanenhaus 2008): less predictable words show greater acoustic prominence (e.g., longer duration, higher F0, less incidence of reduction). We can characterize predictable words in terms of a speech processing model as words whose lexical units are strongly activated due to local priming or frequency in the language (Goldinger, Luce and Pisoni 1989; Luce and Pisoni 1998; Marslen-Wilson 1990). Increased activation levels of lexical units facilitate processing both in perception (e.g., Grossberg 2003; Vitevitch and Luce 1999) and production (e.g., Levelt, Roelofs and Meyer 1999; Marslen-Wilson 1990). To put it simply, a word is perceived and produced more rapidly under increased activation of lexical and sub-lexical units, which occurs when a word is predicted from local context, or for high frequency words, which have high resting activation levels. This facilitation effect is reflected in production in shorter phonetic durations for predictable words, presumably because processing for the following word is started sooner, giving rise to reduced word forms. Words that are less predictable lack this facilitation, and so may exhibit the full duration that is expected on the basis of the lexically specified phonological content of the word. In perception, increased activation of lexical and sub-lexical units results in faster response times in tasks involving lexical access, such

as word recognition. When lexical access is facilitated through high activation levels, there are fewer demands on the processing resources used in speech understanding. We propose that the listener's perception of prominence may directly reflect the demands of speech processing. A listener may judge a word as prominent when processing the word is resource-intensive. Processing demands will be higher, requiring more time, when there are lower activation levels for the lemma or word-form units of the target word, such as with low-frequency, unfamiliar, or otherwise unpredictable words.

In this processing-based account, prominence is both a speaker-based and listener-based phenomenon. Acoustic prominence can arise through lexical access in production, as a speaker-based phenomenon of prosody, resulting in greater acoustic prominence for low-frequency words. This lexical processing effect on acoustic prominence may be at least partly independent of acoustic prominence that reflects phonological structure, e.g., pitch prominence that expresses a phonological pitch accent assigned to a metrically strong syllable as a feature marking pragmatic focus. In addition, perceived prominence can also arise through the processing demands of comprehending speech, which also involves lexical access, as a listener-based phenomenon. Very often, these two sources will converge and a word that is produced with acoustic prominence will also be perceived by the listener as prominent. But the model also allows for cases where a listener perceives a word as prominent, reflecting resource-intensive processing, even when the speaker has not produced the word with strong acoustic cues to prominence. In this account, prominence perception is signal-driven to the extent that speakers' productions contain acoustic cues to prominence, and listeners are sensitive to those cues. But prominence perception is also driven, at least in part, by activation patterns that characterize the listener's expectations in the course of speech processing.

7. Conclusion

This study has shown that untrained listeners reliably perceive prosodic prominence in spontaneous speech, based only on their impressions from real-time listening, and their transcriptions are in agreement well above chance levels. The perceived prominence of a word, as measured through a probabilistic prominence score, is strongly correlated with acoustic measures of prominence taken from the stressed vowel(s), and especially with vowel duration. Thus, prominence perception is partly *signal-driven*. In addition, two factors related to a word's information status – word frequency and repetition in discourse – are also correlated with perceived prominence, providing evidence that prominence perception is also partly *expectation-driven*.

When we examine the acoustic correlates of word frequency (a stronger predictor of perceived prominence than repetition), we find that low-frequency words have a somewhat different set of acoustic correlates than do words that are perceived as prominent by ordinary listeners. In particular, low-frequency words

display spectral emphasis in the high frequency region (a measure of increased vocal effort), while prominent words are notable primarily in their tendency to have longer duration than non-prominent words. This finding, that the acoustic correlates of perceived prominence and word frequency are somewhat different, also points to the conclusion that the relationship between perceived prominence and word frequency is not wholly mediated through the acoustic signal. While listeners may judge a word as prominent based on its acoustic properties, it appears that listeners' prominence rating may also directly reflect word frequency.

Further evidence that a word's information status, as measured through word frequency and the repetition index, influences prominence perception is obtained from hierarchical linear regression models showing that word information measures and acoustic measures of duration and intensity are independent factors that contribute to P-score variance, with the information measures being stronger predictors of P-scores than the acoustic measures combined. As already noted, this finding is evidence that prominence perception is partly expectation driven. In the processing model proposed here, prior experience, either in the context of the discourse or overall experience with the language, facilitates processing and demands fewer resources in the task of recognizing the word. This ease of processing then influences the listener's judgment that the word is not prominent.

From the listener's perspective, a word may be judged prominent because (i) it exhibits enhanced acoustic properties, or (ii) it was relatively unpredicted and thus demanded extra processing resources. Prominence in the first sense is speaker-based and signal-driven, while in the second sense it is listener-based and expectation-driven. And though these two notions of prominence differ somewhat in their acoustic correlates, it's possible that they share a common basis in attention, conceived here as a processing resource (Anderson 2004). We have already described how low-frequency words may require more processing resources than high-frequency words. If we consider processing resources as a form of attention, then we can point to a parallel between the two notions of prominence. A word with acoustic prominence attracts the listener's attention in direct response to the acoustic modulation, while processing a low frequency word demands greater attention because of the lower activation levels of its lexical and/or sub-lexical units. Viewing the results of the present study in these terms, we can say that prominence ratings produced in the task of online prosody transcription reflect the relative attention the listener commits to processing each word in its given discourse context. This notion of prominence perception that relates to attentional resources in speech processing successfully links the speaker and the listener in the communication of prosody.

Acknowledgments

This work is supported by NSF award IIS 07-03624 to Cole and Hasegawa-Johnson. For their varied contributions to the work presented here we thank John

Coleman, Ben Munson, Bob McMurray, Eun-Kyung Lee, Margaret Fleck and members of the Illinois Prosody-ASR research group. Statements in this paper reflect the opinions and conclusions of the authors, and are not endorsed by the NSF or the University of Illinois.

Correspondence e-mail address: jscole@illinois.edu

Notes

1. As noted by a reviewer, the finding that trained listeners assign prominence to a word that exhibits reduced acoustic prominence may reflect the fact that prosody judgments of a trained listener working in a ToBI framework (among others) are based on a richer set of cues, including the visible pitch contour and other aspects of the visual speech display. Of course, visual cues from the speech display do not play a role in ordinary speech communication.
2. Additional findings from acoustic analysis of this dataset are reported in Mo's doctoral thesis in progress.
3. We think that the failure to find a correlation between F0 and perceived prominence may be due in part to the "paradigmatic" normalization method we used, where acoustic measures are normalized within vowel phoneme category, pooling data from all speakers in the corpus. In our ongoing work we are looking at syntagmatic normalization methods, that normalize acoustic measures locally within a stretch of speech defined temporally or on the basis of phonological structure. Preliminary results show some positive correlations between F0 and perceived prominence, but suggest that F0 is not as reliably correlated with perceived prominence as duration or intensity, consistent with the findings of Kochanski et al. (2005).
4. This claim is advanced by Greenburg, Chang and Hitchcock (2001) in their analysis of stress-accent in the Switchboard corpus, although they argue for an effect of vowel height (non-high vowels are more likely to be perceived as stress-accented compared to high vowels), which is not the pattern we observe with the Buckeye data.
5. Word frequency statistics for Switchboard were supplied by Margaret Fleck (p.c.).
6. There are more words in the datasets used for the correlation analyses with log-frequency and P-scores, compared to the datasets used for the analyses of correlation with Repetition, but this difference in sample size is not responsible for the stronger correlations with word frequency. We ran correlations between log-frequency and P-scores for the smaller datasets used in the Repetition analysis and found the same pattern. The linear regression model with log-frequency as a predictor of P-score yields r^2 values of .22 (short excerpts) and .17 (long excerpts), with function words included, while the regression coefficient with repetition as the predictor variable yields r^2 of .01 (short excerpts) and .02 (long excerpts).
7. These findings are further corroborated in correlation analyses with all vowels grouped together ($r = 0.271$). The strength of duration as a primary correlate of P-scores is also observed for duration measures normalized in a local window of five stressed syllables (Mo et al. 2009).

References

- Anderson, John R. 2004. *Cognitive psychology and its implications (6th ed.)*. New York, NY: Worth Publishers.
- Arnold, Jennifer E. 2008. THE BACON not the bacon: How children and adults understand accented and unaccented noun phrases. *Cognition* 108(1). 69–99.

- Aylett, Matthew & Alice Turk. 2004. The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Language and Speech* 47(1). 31–56.
- Bard, Ellen G. & Matthew P. Aylett. 1999. The dissociation of deaccenting, givenness and syntactic role in spontaneous speech. In J. Ohala, Y. Hasegawa, M. Ohala, D. Granville, and A. C. Bailey (eds.), *Proceedings of the XIVth International Congress of Phonetic Science*, 1753–1756. Berkeley, CA: Univ. of California at Berkeley.
- Beckman, Mary E. 1986. *Stress and non-stress accent* (Netherlands Phonetic Archives Series). Dordrecht, The Netherlands: Foris.
- Beckman, Mary E. 1996. The parsing of prosody. *Language and Cognitive Processes* 11. 17–68.
- Beckman, Mary E. & Jan Edwards. 1994. Articulatory evidence for differentiating stress categories. In Patricia A. Keating (ed.), *Papers in Laboratory Phonology III: Phonological Structure and Phonetic Form*, 7–33. Cambridge: Cambridge University Press.
- Bell, Alan, Daniel Jurafsky, Eric Fosler-Lussier, Cynthia Girand, Michelle Gregory & Daniel Gildea. 2003. Effects of disfluencies, predictability, and utterance position on word form variation in English conversation. *Journal of the Acoustical Society of America* 113(2). 1001–1024.
- Bolinger, Dwight L. 1986. *Intonation and its parts: Melody in spoken English*. Palo Alto, CA: Stanford University Press.
- Buhmann, Jeska, Johanneke Caspers, Vincent J. van Heuven, Heleen Hoekstra, Jean-Pierre Martens & Marc Swerts. 2002. Annotation of prominent words, prosodic boundaries and segmental lengthening by non-expert transcribers in the Spoken Dutch Corpus. In *Proceedings of LREC 2002*, Spain, Las Palmas, 779–785.
- Bybee, Joan. 2001. *Phonology and language use*. (Cambridge Studies in Linguistics 94). Cambridge UK: Cambridge University Press.
- Calhoun, Sasha. 2006. Information structure and the prosodic structure of English: A probabilistic relationship. University of Edinburgh PhD thesis.
- Cambier-Langeveld, Tina & Alice Turk. 1999. A cross-linguistic study of accentual lengthening: Dutch vs. English. *Journal of Phonetics* 27. 255–280.
- Cole, Jennifer, Heejin Kim, Hansook Choi & Mark Hasegawa-Johnson. 2007. Prosodic effects on acoustic cues to stop voicing and place of articulation: Evidence from Radio News speech. *Journal of Phonetics* 35. 180–209.
- Dahan, Delphine, Michael K. Tanenhaus & Craig G. Chambers. 2002. Accent and reference resolution in spoken language comprehension. *Journal of Memory and Language* 47. 292–314.
- Fleiss, Joseph L. 1971. Measuring nominal scale agreement among many raters. *Psychological Bulletin* 76(5). 378–382.
- Fosler-Lussier, Eric & Nelson Morgan. 1999. Effects of speaking rate and word frequency on pronunciations in conversational speech. *Speech Communication* 29. 137–158.
- Fowler, Carol A. & Jonathan Housum. 1987. Talkers' signaling of "new" and "old" words in speech and listeners' perception and use of the distinction. *Journal of Memory and Language* 26. 489–504.
- Godfrey, John J., Edward C. Holliman, & J. McDaniel. 1992. SWITCHBOARD: Telephone speech corpus for research and development. In *Proceedings of the IEEE International Conference on the Acoustics, Speech and Signal Processing (ICASP)*, 517–520. San Francisco.
- Goldinger, Stephen, Paul Luce & David Pisoni. 1989. Priming lexical neighbors of spoken words: Effects of competition and inhibition. *Journal of Memory and Language* 28. 501–518.
- Greenberg, Steven. 1999. Speaking in shorthand – A syllable-centric perspective for understanding pronunciation variation. *Speech Communication* 29. 159–176.
- Greenberg, Steven, Shawn Chang & Leah Hitchcock. 2001. The relation between stress accent and vocalic identity in spontaneous American English discourse. In *Prosody in Speech Recognition and Understanding*, 51–56. Proceedings of the ISCA Tutorial and Research Workshop (ITRW), Molly Pitcher Inn, Red Bank, NJ, USA. Accessed from the ISCA Archive, http://www.isca-speech.org/archive/prosody_2001.

- Greenberg, Steven & Eric Fosler-Lussier. 2000. The uninvited guest: Information's role in guiding the production of spontaneous speech. In *Proceedings of the Crest Workshop on Models of Speech Production: Motor Planning and Articulatory Modeling*, Kloster Seeon, Germany, 129–132.
- Gregory, Michelle. 2002. Linguistic Informativeness and speech reduction: An investigation of contextual and discourse pragmatic effects on phonological variation. University of Colorado at Boulder PhD thesis.
- Grossberg, Stephen. 2003. Resonant neural dynamics of speech perception. *Journal of Phonetics* 31. 423–445.
- Heldner, Mattias. 2003. On the reliability of overall intensity and spectral emphasis as acoustic correlates of focal accents in Swedish. *Journal of Phonetics* 31. 39–62.
- Huddleston, Rodney & Geoffrey K. Pullum. 2002. *The Cambridge grammar of the English language*. Cambridge: Cambridge University Press.
- Ito, Kiwako & Shari Speer. 2008. Anticipatory effects of intonation: Eye movements during instructed visual search. *Journal of Memory and Language* 58(2). 541–573.
- Ito, Kiwako, Shari R. Speer & Mary E. Beckman. 2004. Informational Status and Pitch Accent Distribution in Spontaneous Dialogues in English. In Bernard Bel & Isabelle Marlien (eds.), *Proceedings of Speech Prosody 2004*, Nara, Japan, 279–282. Accessed from the ISCA Archive, <http://www.isca-speech.org/archive/sp2004>.
- Kochanski, Greg, Esther Grabe, John Coleman & Burton Rosner. 2005. Loudness predicts prominence: Fundamental frequency lends little. *Journal of the Acoustical Society of America* 118(2). 1038–1054.
- Kuperman, Victor, Mirjam Ernestus & Harald Baayen. 2008. Frequency distributions of uniphones, diphones, and triphones in spontaneous speech. *Journal of the Acoustical Society of America* 124(6). 3897–3908.
- Ladd, Robert D. 1996. *Intonational phonology*. Cambridge: Cambridge University Press.
- Levelt, Willem J. M., Ardi Roelofs & Antje S. Meyer. 1999. A theory of lexical access in speech production. *Behavioral and Brain Sciences* 22. 1–75.
- Luce, Paul A. & David B. Pisoni. 1998. Recognizing spoken words: The Neighborhood Activation Model. *Ear & Hearing* 19(1). 1–36.
- Marslen-Wilson, William. 1990. Activation, competition, and frequency in lexical access. In Gerry T. M. Altmann (ed.), *Cognitive models of speech processing: Psycholinguistic and computational perspectives*, 148–172. Cambridge, MA: MIT Press.
- Mo, Yoonsook. 2008. Acoustic correlates of prosodic prominence for naïve listeners of American English. *Proceedings of the 34th Annual Meeting of the Berkeley Linguistic Society*. Berkeley, CA: Berkeley Linguistic Society.
- Mo, Yoonsook, Jennifer Cole & Eun-Kyung Lee. 2008. Naïve listeners' prominence and boundary perception. In Plinio A. Barbosa, Sandra Madureira & Cesar Reis (eds.), *Proceedings of Speech Prosody 2008*, Campinas, Brazil, 735–738. Accessed from the ISCA Archive, <http://www.isca-speech.org/archive/sp2008>.
- Mo, Yoonsook, Jennifer Cole & Mark Hasegawa-Johnson. 2009. How do ordinary listeners perceive prosodic prominence? Syntagmatic vs. Paradigmatic comparison Poster presented at the 157th Meeting of the Acoustical Society of America, Portland, Oregon.
- Munson, Benjamin. 2007. Lexical access, lexical representation, and vowel production. In Jennifer Cole and José I. Hualde (eds.), *Laboratory Phonology 9*, 201–228. New York and Berlin: Mouton de Gruyter.
- Pierrehumbert, Janet B. 1980. *The phonology and phonetics of English intonation*. MIT, Cambridge, MA PhD thesis.
- Pitt, Mark A., Laura Dilley, Keith Johnson, Scott Kiesling, William Raymond, Elizabeth Hume & Eric Fosler-Lussier. 2007. Buckeye Corpus of Conversational Speech (2nd release) [www.buckeyecorpus.osu.edu] Columbus, OH: Department of Psychology, Ohio State University (Distributor).
- Schwarzschild, Roger. 1999. Givenness, AVOIDF and other constraints on the placement of accent. *Natural Language Semantics* 7. 141–177.

- Selkirk, Elizabeth O. 1996. Sentence prosody: Intonation, stress and phrasing. In John Goldsmith (ed.), *The handbook of phonological theory*, 550–569. Cambridge, Mass.: Blackwell.
- Sluijter, Agaath M. C. & Vincent J. van Heuven. 1996. Spectral balance as an acoustic correlate of linguistic stress. *Journal of the Acoustical Society of America* 100(4). 2471–2485.
- Streefkerk, Barbartje M., Louis C. W. Pols & Louis F. M. ten Bosch. 1997. Prominence in read aloud sentences, as marked by listeners and classified automatically. In R. J. J. H. van Son (ed.), *Proceedings of the Institute of Phonetic Sciences, University of Amsterdam* 21. 101–116.
- Streefkerk, Barbartje M., Louis C. W. Pols & Louis F. M. ten Bosch. 1998. Automatic detection of prominence (as defined by listeners' judgements) in read aloud Dutch sentences. In *Proceedings of the Fifth International Conference on Spoken Language Processing*, Sydney, Australia, 683–686.
- Swerts, Marc. 1997. Prosodic features at discourse boundaries of different strength. *Journal of the Acoustical Society of America* 101. 514–521.
- Tamburini, Fabio. 2005. Automatic prominence identification and prosodic typology, *Proceedings of Interspeech 2005*, 1813–1816. Lisbon.
- Turk, Alice E. & Laurence White. 1999. Structural influences on accentual lengthening in English. *Journal of Phonetics* 27(2). 171–206.
- Vitevitch, Michael S. & Paul A. Luce. 1999. Probabilistic phonotactics and neighborhood activation in spoken word recognition. *Journal of Memory and Language* 40. 374–408.
- Vitevitch, Michael S., Paul A. Luce, David B. Pisoni & Edward T. Auer. 1999. Phonotactics, neighborhood activation and lexical access for spoken words. *Brain and Language* 68. 306–311.
- Watson, Duane G., Jennifer E. Arnold & Michael K. Tanenhaus. 2008. Tic Tac TOE: Effects of predictability and importance on acoustic prominence in language production. *Cognition* 106(3). 1548–1557.
- Wright, Richard. 2003. Factors of lexical competition in vowel articulation. In John Local, Richard Ogden & Rosalind Temple (eds.), *Phonetic interpretation: Papers in Laboratory Phonology VI*, 75–87. Cambridge: Cambridge University Press.
- Yoon, Tae-Jin. 2007. *A Predictive Model of Prosody through Grammatical Interface: A Computational Approach*. University of Illinois at Urbana-Champaign PhD thesis.