To appear in 2012 in G. Elordieta and P. Prieto (Eds.) Prosody and meaning (Trends in Linguistics). Berlin: Mouton de Gruyter.

Information structural expectations in the perception of prosodic prominence

Jason Bishop

University of California, Los Angeles

j.bishop@ucla.edu

Abstract

The present study examined English listeners' knowledge of how the size of a focus constituent is expressed prosodically. English-speaking listeners participated in prominence-rating experiments in which they heard the same production of a simple SVO declarative answer sentence in different focus (i.e., different question) contexts. It was found that when the object was narrowly focused, it was heard as more prominent (and the preceding verb as less prominent) than when that same object was part of broader VP or Sentence foci. This was true regardless of whether the focus was explicitly contrastive. The results are interpreted as the consequence of listeners' experience-based knowledge about how speakers use prosody to express this information structural contrast.

1. Introduction

A recurrent finding in speech perception research is that what a listener "hears" in the acoustic signal is partially influenced by her knowledge of the utterance's linguistic structure. A clear and well-known demonstration of this is the 'phoneme restoration effect' (Warren 1970; Samuel 1981), where listeners are shown to use their knowledge of a word's phonological structure to perceive phonetic information that has been masked or entirely removed from the signal. A similar restorative-like finding is 'perceptual epenthesis' (Dupoux et al.1999), which suggests listeners hear (or do not hear) the presence of segmental information depending on their knowledge of what constitutes a possible phonotactic sequence in their language. Such "top-down" perceptual phenomena are informative because they reveal listeners' expectations; those expectations, in turn, expose details regarding the linguistic structure listeners assign to an utterance.

In the present study, we considered a somewhat less well-investigated part of listeners' expectations, namely those about prosodic rather than segmental information. We probed English-speaking listeners for expectations about how prosodic prominence relates to a certain information structural contrast: the size of the focus constituent (ranging from broad to narrow focus). The prosodic realization of this contrast has been of considerable interest for some time, although there exists much uncertainty about what, if any, conventionalized knowledge speakers and listeners share. Our goals are, first, to highlight a pattern found in a number of recent production studies: an object under narrow focus in an SVO construction is produced with greater acoustic prominence *relative to prenuclear material* than it is when embedded in a broader focus. Then, we attempt to probe listeners for expectations that we can relate, in a detailed sort of way, to that pattern. We find evidence for exactly such expectations, by way of their top-down influence on prominence perception.

The rest of the paper is organized as follows. In Section 2 we describe in detail the information structural contrasts of interest, aspects of the acoustic signal that may correlate with them, and, finally, some previous studies that have probed listeners' knowledge of these correlations. In Section 3 we review recent studies suggesting that listeners' perception of prosodic prominence, like their perception of segmental information, to some extent reflects

their expectations; that is, linguistic knowledge has a top-down influence on prominence perception. In Section 4 two novel experiments explore listeners' expectations about patterns of prominence in broad and narrow focus constructions in English. Section 5 presents a general discussion of the results and their implications; concluding remarks follow in Section 6.

2. The size and type of a focus constituent

Focus is an aspect of a sentence's information structure, and is construed here as the information the speaker presents as *semantically or pragmatically* prominent. There are at least two dimensions along which focus may vary: *size* and *type* (e.g. Gussenhoven 2008; Krifka 2008). The size of a focus constituent[1] is the syntactic constituent under focus; which constituent that will be depends on what information the speaker regards as informative, which itself is dependent on the discourse context in which the sentence is uttered. For example the size of the focus constituent in (1d) will depend on which of the contexts in (1a-c) it is uttered in:

(1) a. *What happened?*
b. *What did you do?*
c. *What did you buy?*
d. I bought a motorcycle.

In response to (1a), the focus constituent of (1d) is the entire sentence, referred to here as *sentence focus*; in the context of (1b), only the verb phrase in (1d) forms the focus constituent, referred to here as *VP focus*. Finally, if (1d) is uttered as a response to (1c), the focus constituent is the object noun-phrase, that is, *object focus*. Where the size of the focus constituent is larger, such as sentence or VP focus, we will refer to them as cases of *broad* focus (Ladd 1980; Selkirk 1984), compared to when the focus is comparatively *narrow* on a single word, such as in object focus.

Regarding focus *type*, we can distinguish two potentially different categories: one which is non-contrastive and one which is contrastive.[2] For the purposes here, *non-contrastive focus* will be used to refer to cases like that in (1), namely, the information which is required by a WH-question. In such cases, there is no explicitly mentioned alternative to the answer in the discourse. *Contrastive focus* will be reserved for cases where the focus of a sentence does have an explicitly mentioned alternative (often referred to as "corrective focus"), as in *motorcycle* in (2b):

(2) a. *Did you buy a car?*
b. (No) I bought a motorcycle.

Note that when cases of contrastive focus, as defined here, are discussed, they are often also cases of narrow focus (e.g. Baumann et al. 2008; Hanssen et al. 2008). However, focus size and focus type are (in principle) orthogonal properties of information structure, and thus it is possible for a focus of any size to either have or not have an explicitly mentioned alternative in the discourse. In the present case, (2b) could have contrastive VP focus if uttered in response to a question such as "*Did you work today?*".

A matter of much recent interest is the relation these two dimensions of focus have to prosodic realization. Regarding the size of the focus in the simple subject-verb-object (SVO) constructions considered in this paper, one account, the Focus Projection Hypothesis (Selkirk 1995; see also Gussenhoven 1984) predicts an ambiguity. According to this hypothesis, a single pitch accent on the object is said to be appropriate for the answer to each of the questions in (1), and so a distinctly narrow or distinctly broad focus meaning of the sentence

is not expected to be expressible prosodically. However, it has been known for some time that speakers do not necessarily pronounce foci of different sizes equivalently. For example, Ladd (1996) discussed the possibility for English speakers to disambiguate narrow focus on an object from broad VP focus by placing more "emphasis" on the focused object – pronouncing it, for example, with a higher f0 peak. Further, Gussenhoven (1983) found that listeners auditorily detected differences in sentences produced in a broad or narrow focus context; when asked to judge the prominence of verbs, listeners rated those produced in broad focus contexts as more prominent than those produced in narrow object focus contexts. Since listeners heard such sentences separately from their contexts, this must have been the result of cues that speakers in Gussenhoven's study encoded in the sentences.

Subsequent to some of the earlier observations, controlled production studies have suggested a rather systematic use of acoustic features to distinguish focus constituents of different sizes. It must be emphasized that this evidence comes from West Germanic languages, and we restrict our discussion to these particular languages.[3] In Dutch, for example, it has been shown that narrow focus is distinguishable from broad focus with respect to segmental durations (longer in an object under narrow focus) and the shape of the nuclear accent on the object (a steeper fall under narrow focus) (Hanssen et al. 2008; see also He et al. 2011). For German, Baumann, Grice and Steindamm (2006; see also Baumann et al. 2008) found that as focus narrowed from broad sentence focus to narrow object focus, the nuclear accented object was pronounced with longer segmental durations, with a higher f0 peak relative to prenuclear accents, and the probability of prenuclear accents on verbs decreased. Most important for our discussion is that studies that have considered the phonetic realization of objects in relation to surrounding material have reported the same basic patterns for English speakers (Eady et al. 1986; Sityaev and House 2003; Xu and Xu 2005; Löfstedt 2006; Jun 2008; Breen et al. 2010). One of the studies that stands out particularly is that by Breen et al. (2010), for two reasons. First, an elicitation method was used in which speakers gave answers to various focus questions based on pictures rather than read materials. Second, an additional experiment put speakers in a highly communicative situation where the goal was to purposefully disambiguate between broad and narrow focus. In both cases, speakers' productions (using a number of acoustic parameters) were correctly classified for focus size by a statistical model, somewhat more successfully when speakers were aware of the ambiguity.

Fewer studies have examined the prosodic realization of focus type, and among those that have, the results are somewhat less uniform compared to what has been reported for focus size. While two of the studies mentioned above, Hanssen et al. (2008) and Baumann et al. (2008), found no differences in speakers' productions of narrow contrastive and narrow non-contrastive focus, some studies with English speakers have. Bartels and Kingston (1994), for example, claim that contrastive focus is most reliably distinguished from non-contrastive focus by the height of the accent peak, f0 being higher for contrastive focus. Consistent with this, Ito, Speer and Beckman (2004) found speakers more likely to use a prominent pitch excursion (the ToBI L+H* rather than the H*) to mark foci as contrastive. Finally, Breen et al. (2010) reported finding acoustic differences between productions of contrastive and non-contrastive focus when speakers were deliberately trying to communicate the distinction and were given feedback on their success. Although their findings do not support those of the previous two studies of English in terms of f0's role (indeed they find the opposite), Breen and colleagues found that when speakers intentionally tried to express contrastive focus, they did so using greater intensity relative to surrounding words.

Having considered recent investigations into the phonetic realization of focus size and focus type, two possible generalizations can be made regarding how speakers might encode these aspects of information structure prosodically. The first is that speakers tend to use

increased acoustic prominence on a sentence-final object when it is narrowly focused compared to when it is situated within a larger focus constituent. The second, similarly, is that speakers produce such an object with more acoustic prominence if it is being used contrastively. It must also be noted that these generalizations might be more applicable to a certain kind of speech, that which is highly communicative or explicitly listener-directed. What these basic observations would lead us to predict, is that listeners will exhibit some corresponding expectations when faced with the task of interpreting the information structure intended by the speaker.

In fact, listeners' have been shown to be sensitive to prosody in this regard, but not all experimental measures seem to reveal this. For example, although Gussenhoven (1983) found that productions of broad and narrow focus were auditorily distinguishable, he found no evidence that the differences were useful in a task requiring listeners to match answer sentences with the correct question contexts. This basic result has been replicated in more recent studies using similar methodology. Birch and Clifton (1995) and Welby (2003), for example, found that listeners rated an answer sentence with a nuclear accented object and optionally prenuclear accented verb as equally "appropriate" in broad VP or narrow object focus contexts. Evidence that there is more to the matter than this, however, comes from one of these very studies; while Birch and Clifton's experiment found that appropriateness decisions did not disambiguate prosody–focus size pairings, their reaction times for making those decisions did. They found that response times for rating appropriateness were somewhat slowed when a sentence under VP focus lacked a prenuclear accent on the verb. This interesting finding suggests that, at least in processing, listeners exhibited a preference for a particular broad focus pronunciation (that is, one in which the verb was more prominent) that was not detectable in their judgments of felicity.

Rump and Collier (1996) present evidence that further highlights the importance of the kind of task listeners are faced with. These authors report on experiments in which Dutch listeners were presented with the standard SVO sentences that followed questions that asked for (among other foci) broad sentence or narrow object focus. In the first experiment, the task for listeners was to adjust the f0 height of a prenuclear (on the subject) and a nuclear (on the object) accent by choosing from a series of previously synthesized stimuli; in a second experiment, listeners were to match sentences with prenuclear and nuclear accents of various relative heights with question sentences that asked for a broad or narrow focus. In both experiments, listeners showed very clear preferences for a relatively higher peak on the object when it was narrowly focused compared to when it was part of a broad sentence focus, consistent patterns reported in production studies. Thus there seems to be something about the task of judging acceptability of pronunciations that elicits a more generous response from listeners, and fails to detect their expectations about how prosody can disambiguate. This might be interpreted as an indication that the relevant phonetic cues have little actual communicative value. Yet, the data reported by Breen et al. (2010) suggest such a claim is too strong. Their second experiment was similar to Gussenhoven's (1983) experiment, in that listeners were to match a speaker's answer productions with the "correct" question contexts. Unlike Gussenhoven's earlier study, however, Breen and colleagues found that listeners were able to do this with considerable success, rarely (only 13% of the time at most) confusing intended narrow object focus with intended broad focus productions. The important difference might be that, again, stimuli in Breen and colleagues' experiments were produced by speakers whose goal was to disambiguate. The fact that listeners (not only the statistical model) were able to distinguish the meaning from this kind of speech is not a trivial result. It provides us with evidence that listeners have knowledge – useful knowledge – about how broad and narrow focus productions should differ; the cues speakers gave with the intention of disambiguating were effective in leading listeners to recover the intended meaning.

Evidence that listeners have similar expectations about the prosodic realizations of focus *type* is also available, and can be found in Ito and Speer (2008). In their eye-tracking study, listeners were shown to respond differently to prominent accent peaks (labeled L+H* by ToBI-trained transcribers) than to less prominent accent peaks (labeled H*), such that the *more* prominent accent peaks evoked a contrastive interpretation. Using the visual world paradigm, listeners in their study were presented with an array of ornaments of different colors and were given instructions for hanging them on a Christmas tree. Ito and Speer found that when asked to hang, for example, a *blue drum* ornament, with a more prominent accent on the adjective *blue* in a context such as "*Hang the green drum…now hang the BLUE drum"* elicited a contrastive interpretation, indicated by listeners' fixations on suitable alternatives. This was far less likely to happen if the adjective carried the *less* prominent accent (e.g. *"Hang the green drum…now hang the blue drum"*). Indeed an infelicitous use of a prominent accent on an adjective (e.g. *"Hang the red angel… now hang the BLUE drum"*) evoked contrastive responses from listeners, causing them to fixate on alternatives to the previously mentioned ornament type, even while hearing conflicting segmental information. Thus, listeners show expectations regarding the prosodic realization of focus type, much like they do for focus size. In the case of focus type, the expectation seems to be that contrastive foci are more prominent than non-contrastive foci.

We have so far discussed a considerable amount of evidence bearing on how the meaning distinctions of focus size and focus type may correspond to prosodic realization. In particular, it was noted that both narrow focus and contrastive focus seem to correlate with greater acoustic prominence (relative to adjacent material in the utterance) compared with broad focus and non-contrastive focus. Further, we have reviewed some of the available evidence for listeners' *knowledge* of these relationships. Before going on to present two experiments that probe the consequences of this knowledge for the perception of prosodic prominence, it is first necessary to distinguish perceived prominence of prosodic units from acoustic prominence, and briefly discuss how the two might relate.

3. Perceived prominence

Perceived prominence is the listeners' subjective impression of "prosodic strength". Although it is not defined in terms of the acoustic signal, it is well established that a number of acoustic cues serve as predictors of perceived prominence, some of which have been alluded to above in relation to productions of focus size and type. In particular, increased segmental durations, greater intensity (both contributing to "loudness"), and salient aspects of the fundamental frequency (f0) contour – namely, peaks, valleys and movements – are all associated with greater perceived prominence. Although there is some disagreement across experimental studies with respect to their relative importance to English listeners (e.g. Beckman 1986; Gussenhoven et al.1997; Kochanski et al. 2005), much of the variance in listeners' judgments of prominence can be accounted for on the basis of a model that includes some combination of these features. Cole, Mo and Hasegawa-Johnson (2010), for example, found that intensity and duration were strong predictors of the probability that linguistically untrained listeners would judge words in excerpts taken from the Buckeye Corpus (Pitt et al. 2007) as prominent. Findings consistent with this are reported in Kochanski et al. (2005) and Mo (2008). What this would seem to suggest is that a listener's impression of prosodic prominence can be modeled as a *signal-based* (or "bottom-up") process – i.e., the result of acoustic cues.

However, there is also evidence that the perception of prominence emerges from *non-signal-based* ("top-down") factors. In addition to duration and intensity, Cole and colleagues'

study examined correlations between listeners' prominence ratings and lexical and discourse variables. They found that a word's lexical frequency and number of previous occurrences in the experimental materials were negatively correlated with the probability of its being judged as prominent. Although this pattern was consistent with the acoustic prominence of those words (that is, lexical frequency was also negatively correlated with duration and intensity), the relationship between word frequency and prominence judgments was partly independent of the acoustic features.[4] A result similar to Cole and colleagues' was reported in Eriksson, Thunberg and Traunmüller (2001) for Swedish. In their study, the authors compared two types of models of prominence judgments (for sentences produced by multiple speakers): one with only non-signal-based linguistic factors and one with only acoustic predictors. The non-signal-based model of listeners' judgments – which included whether or not a syllable was phonologically capable of carrying an accent, or whether or not the word was used contrastively – accounted for 57% of the variance in listeners' judgments, an improvement over the 48% accounted for by the acoustics-only model. As in Cole and colleagues' data set, the signal-based and non-signal-based variables in this study were surely correlated with each other. However, the higher performance of the non-signal-based model is an interesting and suggestive finding. Results of this type suggest that listeners may be making prominence judgments that are consistent with patterns found in productions generally, but not necessarily in the particular stimulus at hand. The question we now wish to ask is whether listeners' experience with productions of different information structures might influence their impressions of prosodic prominence in a similar manner.

4. Probing listeners' expectations for prominence

The previous section discussed evidence that listeners' perception of prominence can be influenced by factors not directly found in the acoustic information they receive, but could be seen as reflecting their expectations based on experience with speakers' productions. In Section 2 we discussed evidence bearing on what their experience with productions might look like with respect to foci of different sizes and types. For narrow focus on an object, English speakers tend to use greater phonetic prominence on that object compared to when it is situated within a broader VP or sentence focus. A similar pattern is shown for non-contrastive versus contrastive focus. An important aspect of this general pattern, however – one which we might predict listeners to have knowledge of – had to do with how speakers implement this prominence. That is, they tend to do so not just by manipulating prominence on the object directly, but indirectly by suppressing the prominence of surrounding material (such as material in the prenuclear region). In what follows, we present two experiments that were designed to test whether listeners showed any expectation for these patterns in their judgments of prosodic prominence in controlled materials. In particular, we asked whether listeners would judge the same production of a sentence differently depending on the size of its focus constituent. This was tested in English making use of cases such as (1), repeated here as (3):

(3) a. *What happened?*
b. *What did you do?*
c. *What did you buy?*
d. I bought a motorcycle.

The same pronunciation of (3d), with an (optionally) prenuclear accented verb and (obligatorily) nuclear accented final object, is said to be appropriate in each of the contexts in (3a-c) (Gussenhoven 1983; Selkirk 1995; Birch and Clifton 1995; Welby 2003). However, as discussed earlier, the information structure of each differs in terms of the size of the focus

constituent. Based on previous studies of listeners' prominence judgments, and in line with expectations listeners could have about productions, it is predicted that a narrowly focused object ((3d) heard in the context of (3c)), will be judged as more prominent than an object that is part of a broader focus ((3d) heard in the context of (3a) or (3b)). This might be true even in the absence of any acoustic cues that could disambiguate the three contexts. This is tested for non-contrastive focus in Experiment 1. Although focus size and type, as discussed above, vary independently, contrastive focus is known to be accompanied by extra acoustic prominence in speakers' productions, and previous studies have sometimes treated narrow contrastive focus as a separate category (e.g. Baumann et al. 2008; Hanssen et al. 2008). For this reason, Experiment 2 tested the effect for focus size when the focus constituent was explicitly contrastive. In both experiments, untrained listeners were asked to listen to sentences in different contexts, each intended to encourage a different interpretation, and to assign prominence ratings to the verbs and objects in those sentences.

Table 1: Example set of question-answer dialogues used in the Experiment 1.

Focus Condition	Question Context	Answer Sentence
Sentence-Foc	What happened yesterday?	I bought a motorcycle.
VP-Foc	What did you do yesterday?	
Obj-Foc	What did you buy yesterday?	

4.1. Experiment 1

4.1.1. Method

4.1.1.1. Materials

Recorded sets of mini-dialogues, question-answer pairs as in (3), above, were used as experimental materials for a prominence rating task. A dialogue set included one version of an SVO answer sentence such as *I bought a motorcycle*, and three different set-up questions, all WH-questions such as *What happened yesterday?*, *What did you do yesterday?,* or *What did you buy yesterday?* The answer sentence was to be used as the test sentence for which linguistically naïve listeners would provide prominence ratings. The set-up questions allowed for the pragmatic manipulation of the size of the focus in test sentences, and resulted in three experimental conditions: one in which the entire answer sentence was to be interpreted as the focus (sentence focus), one in which the verb-phrase was the focus (VP focus), and one in which only the object was the focus (object focus) (Table 1; for the full list of stimuli used, see the Appendix). The dialogues were recorded and used to create stimuli as follows. Two native speakers of American English (both linguistics graduate students) read 17 sets of question-answer exchanges from a printed booklet. The printed dialogues contained no intonational annotations and neither the speaker of the questions (a female), nor the speaker of the answers (a male) was instructed on how to produce the sentences, except to read the exchanges as naturally as possible. The speakers were recorded (over two separate channels) while speaking into head-mounted microphones in a sound-attenuated booth in the UCLA Phonetics Lab. Recordings of the speakers' questions and answers were digitized at 22.05 kHz, saved and stored as separate .wav files.

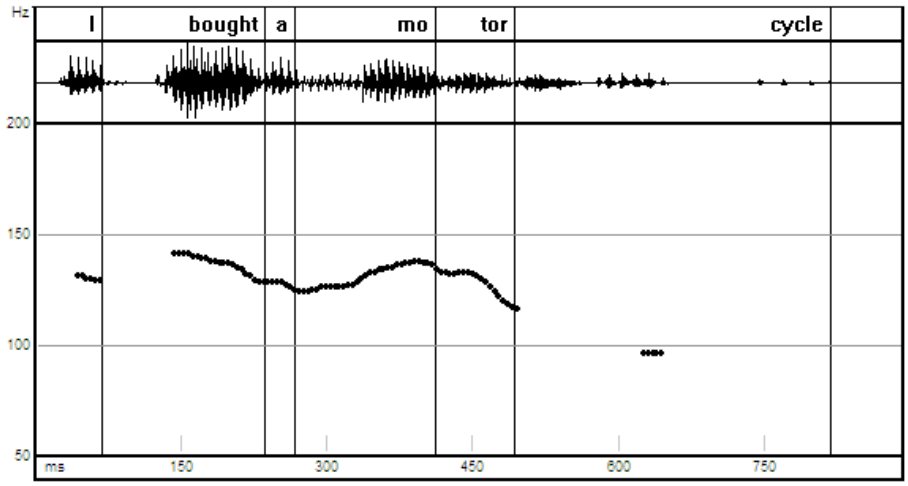

Figure 1: Example waveform and pitch track for the sentence *I bought a motorcycle.*, recorded as answer to the VP focus question *What did you do yesterday?*. This production (ToBI transcribed with H* on both the verb and object) was presented as an answer to each of the questions in the three focus conditions.

Because the purpose of the study was to test the independent effect of information structure on the perception of prominence for words in the answer sentences, it was necessary to hold all acoustic information constant in those sentences across the conditions in which they would be presented. This was accomplished by extracting the recordings of answer sentences produced in response to VP focus questions in the original recordings and using them as the test sentences for all three focus conditions. Thus, for example, the production of the answer sentence *I bought a motorcycle* (Figure 1), originally produced as an answer to the VP focus question *What did you do yesterday?*, was made to follow each of the three different questions recorded in that set. Based on previous research, it was highly expected that each of the VP focus answer sentences to be used as stimuli would be pronounced as a single prosodic phrase with the nuclear accent on the object (Birch and Clifton 1995; Beckman 1996; Welby 2003; Breen et al. 2010). Note, however, that since the only restriction on the speakers' readings of the dialogues was that they were impressionistically natural and felicitous, no control was exerted over how the test sentences might have been produced within those simple limitations. In this sense, no *a priori* assumptions about how speakers should produce the test sentences entered directly into the form of the stimuli. However, some prosodic aspects of an "appropriate" answer to these kinds of questions are not predictable for an individual utterance. As discussed in Section 2, the presence or absence of a prenuclear accent, and also the type of nuclear accent (e.g. Gussenhoven 1983; Selkirk 1995; Ladd 1996; Jun 2008), can vary. It was therefore necessary to identify the intonational pattern of the sentences used as stimuli. Those data are reported here in the form of ToBI (Tones and Break Indices) annotations.[5] Two linguists trained in using the model for Mainstream American English independently transcribed tones (not break indices) for the verb phrase of the 17 answer sentences used as test stimuli. Labeling of test sentences was done without any question context included in the sound file. The tones assigned by the first labeler are shown in Table 2. It is often noted (e.g. Syrdal and McGory 2000; Calhoun 2006; Breen et al. in press) that one of the least reliable distinctions in prosodic transcription of English is that between what in ToBI are represented by H* and L+H*, and for this reason the two categories are often collapsed in calculating transcriber agreement. Here we maintained the distinction, but collapsed the smaller and likely less reliable one between !H* and the L+!H*. Agreement between the two labelers for verbs in the test sentences was

100%, all being transcribed with a prenuclear H*. Agreement for tones on objects was 76.4%. While the two labelers agreed that all of these objects carried a phrase-final pitch accent, the disagreements that arose regarded whether that pitch accent was a H* or !H*, and Table 2 shows the labeler whose tended to use H*.[6] Agreement for boundary tones was 100% (for both intermediate and intonational phrases). The ends of sentences were usually marked by L targets, although in five cases the speaker's productions showed a fall from the nuclear pitch accent (associated with the intermediate phrase), followed by a slight rise (associated with the intonational phrase).

Table 2: Intonational structure of the 17 test sentences, described in the form of ToBI transcriptions.

Verb	Object	ip-Boundary	IP-Boundary	# of items
H*	H*	L-	L%	7
H*	!H*	L-	L%	4
H*	!H*	L-	H%	3
H*	H*	L-	H%	2
H*	L+H*	L-	L%	1

MS PowerPoint presentations were created to present the recorded stimuli. The completed 51 recorded dialogues (17 test sentences, each occurring in three question contexts) were arranged in three different pseudorandomizations, intermixed in each with 37 non-experimental filler dialogues. The fillers closely resembled the experimental dialogues in most respects (including the conditions in which they appeared) but differed from them (a) in terms of syntactic structure (fillers contained adjuncts) and (b) they were subject to additional focus conditions (double focus on subjects and objects, and narrow focus on verbs). In each ordering of the stimuli, members of a crucial set (e.g. Table 1) were separated by at least 6 question-answer dialogue items. The PowerPoint presentations of the stimuli contained only an item number and a play button on each slide, which participants used to listen to the items; neither the orthography nor any visualization of the prosody/acoustics appeared on the screen.

4.1.1.2. Participants

30 native speakers of American English were recruited from the University of California, Los Angeles to participate as listeners in a prominence rating experiment. All were undergraduate students or (non-academic) employees at the university. Many of the student participants were linguistics or psychology majors, although none had any training in intonational phonology or the transcription of prosody. All participants confirmed they had no previous diagnosis of a hearing or communication disorder, and all were paid for their participation.

4.1.1.3. Procedure

Listeners participated in a naïve prominence "transcription" task. Listeners were able to proceed through experimental items in the PowerPoint presentation at their own pace,

listening to each recorded dialogue as many times as they wished, although they were discouraged from listening more than two or three times. As they played each item, they were to listen for how "stressed" words in the male speaker's answers sounded. The experimenter emphasized to participants that their task was to listen to how the answer sentences were pronounced, and this was described to each participant as follows:

> "This experiment is about how speakers pronounce words in a sentence. Your task is to tell us as accurately as possible how stressed the underlined words sound relative to other words in the sentence. By "stressed" we mean "how much did the speaker use his voice to make the word stand out?"

In many experiments involving prominence judgment tasks, the participant is asked to judge which of two words or which of two accent peaks in a sentence is more prominent (e.g. Pierrehumbert 1979; Rietveld and Gussenhoven 1985; Gussenhoven et al. 1997), or to pick out only prominent rather than non-prominent words in a recording (Cole et al. 2010). However, as mentioned above, all the test sentences used here were transcribed pronounced with a nuclear accent on the sentence-final object, and no acoustic manipulation of the stimuli was carried out. Thus, for the stimuli used here, it is highly likely that asking participants to identify the most prominent word in the test sentence would result in 'object' responses in most or nearly all cases. Therefore, a more gradient method of response was used to collect participants' judgments across conditions. Participants were provided with printed transcripts of the dialogues they heard, ordered and numbered as they appeared on the PowerPoint lists. Participants were instructed to follow along on the transcript and to provide ratings of "stress" from 1 ("not at all stressed") to 5 ("very stressed") for words that were underlined on that transcript. These words were the verbs and the objects in the answer sentences, and they were to write in their ratings above the word. An example of how these items appeared on the transcript is shown in (4); the numbers appearing above the underlined words are hypothetical examples of how listeners provided their prominence judgments.

(4) a. Q: What did you do yesterday?
2 4
A: I <u>bought</u> a <u>motorcycle</u>.

b. Q: What did you eat at the picnic?
2 5
A: I <u>ate</u> a <u>hamburger</u>.

Before beginning the experiment, participants completed a short practice session of three dialogue items to familiarize themselves with the style of the dialogues, the speakers and the general set-up for the task. After completing the practice session and asking questions, participants listened to the 51 test and 37 filler dialogues binaurally over Sony MDR-V500 closed, dynamic headphones at a comfortable listening volume (held constant across participants) in a sound attenuated booth in the UCLA Phonetics Laboratory. They provided prominence ratings as above for verbs and objects in each sentence (30 listeners × 17 test sentences × 2 words (verbs and objects) × 3 focus conditions (sentence focus, VP focus, object focus) = 3,060 ratings). These ratings served as the outcome variable in a mixed-effects linear model (using the *lmer* function in R). The predictors in the model included *listener* and *item* as random factors and the following fixed effects factors: the *word* rated (verb or object), the experimental manipulation *focus size* (sentence, VP, or object), and the interaction of these two factors.

4.1.2. Results

Average listener ratings for objects and verbs are shown for each of the focus size conditions in Figure 2. A comparison of the fit of the full model (i.e., the model containing the three fixed effects parameters) to the data was shown to be superior to that of a baseline model containing only random effects parameters (logLik = -4235 vs. logLik = -4244; *anova* function in R: $p < .0001$). Table 3 shows the results of the model when "verb" and "object focus" are the default values (i.e., comparison groups) for *word* and *focus size*, respectively. According to the model, there was a significant effect for *word*; overall, verbs (mean = 2.87, SD = 1.13) were associated with lower prominence ratings than objects (mean = 3.03, SD = 1.09). There was a smaller but significant main effect for the focus manipulation, *focus size* that indicated words were rated as less prominent under the two broad focus conditions (sentence and VP) compared with the narrow object focus condition. Both of these effects, however, are best understood in terms of the significant interaction between *word* and *focus size*. As is easily seen in Figure 2, although objects were judged as more prominent than verbs in each of conditions, the difference was greatest in the object focus condition, due to both object as well as the verb being significantly different from the other two conditions. With respect to the two broad focus conditions, the results of the model indicate they had the same distance from the comparison group, suggesting no significant differences between sentence and VP focus were likely. To confirm this, the default value for *focus size* in the (same) model was reset to "VP focus" so that a direct comparison could be made with "sentence focus"; the groups were statistically the same ($p > .1$).

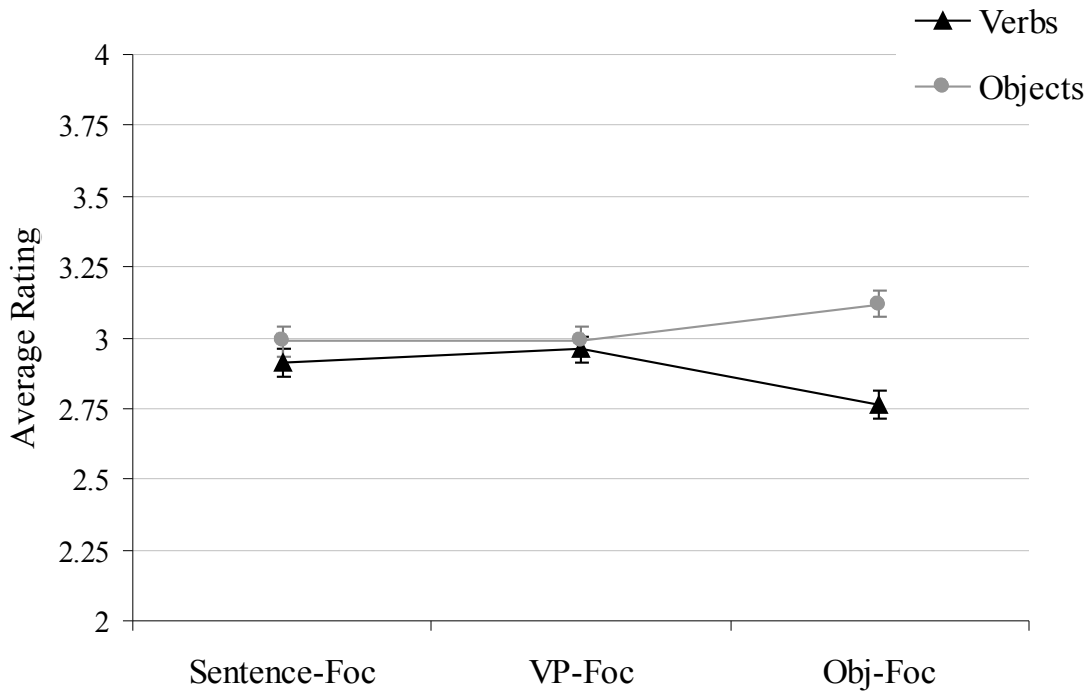

Figure 2: Average prominence ratings for verbs and objects in test sentences in the three focus conditions in Experiment 1. '1' is lowest in prominence, '5' is highest. Error bars show standard error.

Table 3: Results of the linear mixed effects model of listeners' prominence ratings in Experiment 1.

Fixed effects	*Estimate*	*Std. Error*	*t value*	*p value*
(Intercept)	3.1191	0.1201	25.97	< .0001
Word (verb)	-0.3578	0.0590	-6.07	< .0001
Focus Size (Sen)	-0.1304	0.0590	-2.21	.0271
Focus Size (VP)	-0.1304	0.0590	-2.21	.0271
Word*Focus Size (Sen)	0.2824	0.0834	3.39	.0007
Word*Focus Size (VP)	0.3284	0.0834	3.94	< .0001

Random effects	*Variance*	*Std. Deviation*
Subject (Intercept)	0.3527	0.5939
Item (Intercept)	0.0157	0.1254
Residual	0.8871	0.9419

4.1.3. Discussion

The results of Experiment 1 demonstrate two important things. The first is that listeners' judgments of prosodic prominence are significantly and independently affected by their interpretation of the utterances' information structure. More specifically, a nuclear accented object was heard as more prominent when it was narrowly focused rather than when it was situated within a broader focus constituent. The second main point is the specific way in which this effect manifested itself in those judgments; the pattern was such that the object was made prominent relative to the prenuclear verb. The significant interaction between *word* and *focus size* indicated that objects were not simply heard as more prominent, but also that verbs were heard as less prominent; in fact, the lowering effect of prominence on verbs was actually numerically stronger than the increasing effect on objects. It should be noted that this aspect of the pattern is not logically necessary, but, again, it is exactly what we expect if listeners have clear expectations about what speakers do.

An additional point we might make is that, with respect to the effect of focus size, listeners could not have been responding to a something actually in the signal; it was the exact same auditory stimulus (i.e., the declarative SVO answer sentence) they heard in each focus size condition, and only the context (i.e., the preceding question) varied. However, did listeners simply ignore the signal and rely completely on their expectations about information structure? This would seem very unlikely, and indeed seems not to be the case; as indicated by the model, verbs were rated, overall, as less prominent than objects. This is the pattern we would expect from signal-based cues; the objects, rather than the verbs, were nuclear accented and nuclear accented words are generally (but not always) phonetically more prominent than prenuclear accents. Thus the interpretation here is that listeners were attending to the signal, probably primarily to the signal, and their expectations about the realization of the relevant information structural contrast were modulating those judgments. Further evidence for this interpretation will be shown below, in Experiment 2, which, making use of different items and listeners, attempted to replicate the effect of focus size for contrastive rather than non-contrastive focus.

4.2. Experiment 2

4.2.1. Method

4.2.1.1. Materials

A second set of 51 short dialogues were recorded by the same two speakers used in Experiment 1, and were analogous in structure to those in Experiment 1. However, in order to test for the effect of focus size for contrastive focus, the dialogues differed in the following ways. All of the questions read by the female speaker were embedded in complementizer phrases headed by *because*, which were themselves preceded by a set-up question (see Table 4; complete list shown in Appendix). For example the female speaker read questions such as *Why aren't you hungry?* and offered a possible reason which was intended as a set-up to the interpretation of the focus structure of the answer. That proposition was then corrected in the answer sentence read by the male speaker. Thus, in the context of *Why's your wife mad?... because you lost your job?"*, the answer "*No, because I bought a motorcycle."* is assumed to be a correction to the VP, *lost your job*. In the context of "*Why's your wife mad?... because you bought a car?"* that same answer is assumed to be a correction only to the object, *motorcycle*. These materials were recorded as in Experiment 1, and productions of the answer sentence spoken in a VP context (e.g. Figure 3) were saved and paired with the three different questions in the same way. ToBI annotations were again assigned to the test sentences by two labelers. The first labeler's transcriptions are shown in Table 5; agreement for accents on verbs was 76.5%, accents on objects 76.5%, boundary tones 100% (all sentences being transcribed with low boundary tones after the object). Disagreements in
assignments for accents on verbs regarded the presence or absence of an accent and whether an accent, if present, was H* or !H*. Disagreements for objects involved whether they bore a H* rather than L+H*; rate of agreement as to the presence of a nuclear accent on the object, however, was 100%.

Table 4: Example set of stimuli used in Experiment 2.

Focus Condition	Question Context	Answer Sentence
Sentence-Foc	Why's your wife mad… because the roof's leaking?	No, because I bought a motorcycle.
VP-Foc	Why's your wife mad… because you lost your job?	
Obj-Foc	Why's your wife mad… because you bought a car?	

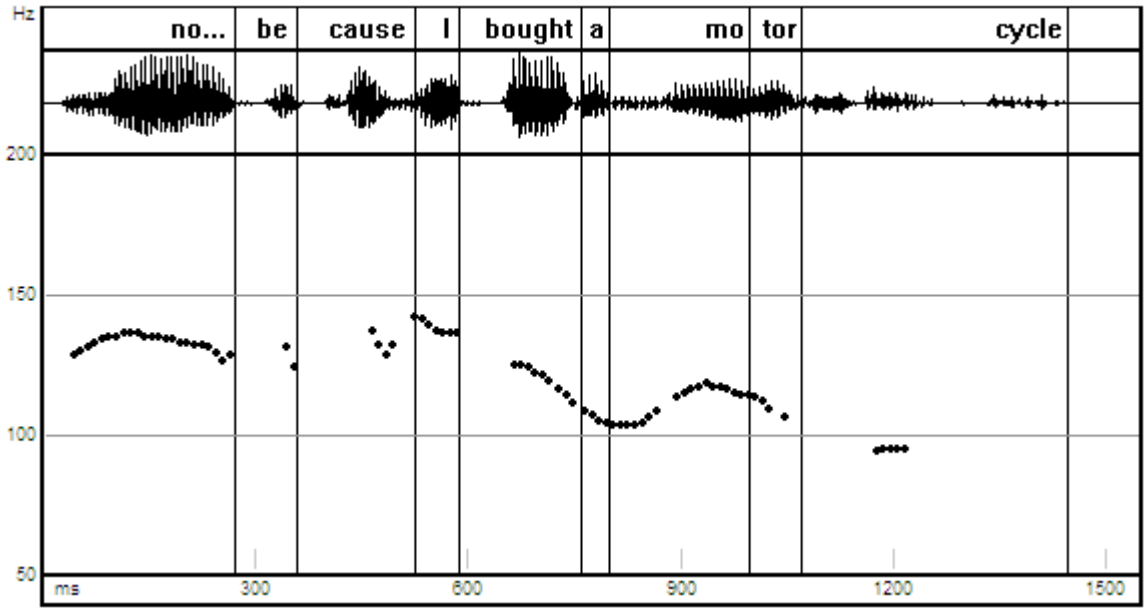

Figure 3: Example waveform and pitch track for the sentence *No, because I bought a motorcycle*, recorded as an answer to the VP focus question *Why's your wife mad…because you lost your job?*. This production (ToBI transcribed with an unaccented verb and L+!H* on the object) was presented as an answer to each of the three focus conditions.

Table 5: Intonational structure of the 17 test sentences, described as ToBI transcriptions. 'Ø' indicates the absence of a pitch accent.

Verb	Object	ip-Boundary	IP-Boundary	# of items
Ø	L+H*	L–	L%	6
H*	H*	L–	L%	3
H*	L+!H*	L–	L%	3
Ø	!H*	L–	L%	2
H*	!H	L–	L%	1
H*	L+H*	L–	L%	1
!H*	L+H*	L–	L%	1

4.2.1.2. Participants

30 native speakers of American English were recruited from the University of California, Los Angeles as in Experiment 1. No participant reported any previous diagnosis or knowledge of a hearing or communication disorder; all were paid for their participation.

4.2.1.3. Procedure

The procedure for Experiment 2 was carried out as for Experiment 1.

4.2.2. Results

It was discovered that two of the participants in Experiment 2 had also participated in Experiment 1 two months prior, and so data from these two participants were removed from the analysis. Average listener ratings across conditions for the remaining twenty-eight subjects are shown in Figure 4. A linear mixed-effects model was fitted as in Experiment 1, using the same parameters. The fit of that full model (logLik = -3594) to the data was significantly better than that of the base-line model containing only random effects factors (logLik = -3810) according to a log likelihood ratio test ($p < .0001$). The outcome of the full model is shown in Table 6 and indicates the following. There was a significant effect for *word*, verbs (mean = 2.69, SD = 1.01) being judged as less prominent than objects (mean = 3.37, SD = 1.15). As in Experiment 1, there was also an effect for *focus size* that indicated words were rated as less prominent under the two broad focus conditions (sentence and VP) compared with the narrow object focus condition. However, also as in Experiment 1, both of these effects must be evaluated in terms of their interaction in the model; the highly significant interaction between *word* and *focus size* indicated that the experimental manipulation did not influence both verbs and objects in the same manner. As is clear in Figure 4, the interaction effect is driven by the difference between verbs and objects in the narrow focus condition; while objects were rated as more prominent when they were interpreted as narrowly focused, the verbs preceding them were perceived as significantly less prominent. To directly compare the two broad focus conditions, the default comparison group was reset as in Experiment 1; this indicated that "VP focus" and "sentence focus" did not differ in their effect on listeners' ratings ($p > .1$).

Table 6: Results of the linear mixed-effects model of listeners' prominence ratings in Experiment 2.

Fixed effects	*Estimate*	*Std. Error*	*t value*	*p value*
(Intercept)	3.4517	0.1465	23.55	< .0001
Word (verb)	-0.8446	0.0535	-15.78	< .0001
Focus Size (Sen)	-0.1324	0.0535	-2.47	.0135
Focus Size (VP)	-0.12	0.0535	-2.20	.0280
Word*Focus Size (Sen)	0.2416	0.0757	3.19	.0014
Word*Focus Size (VP)	0.2710	0.0757	3.58	.0003

Random effects	*Variance*	*Std. Deviation*
Subject (Intercept)	0.5029	0.7092
Item (Intercept)	0.0354	0.1881
Residual	0.6821	0.8259

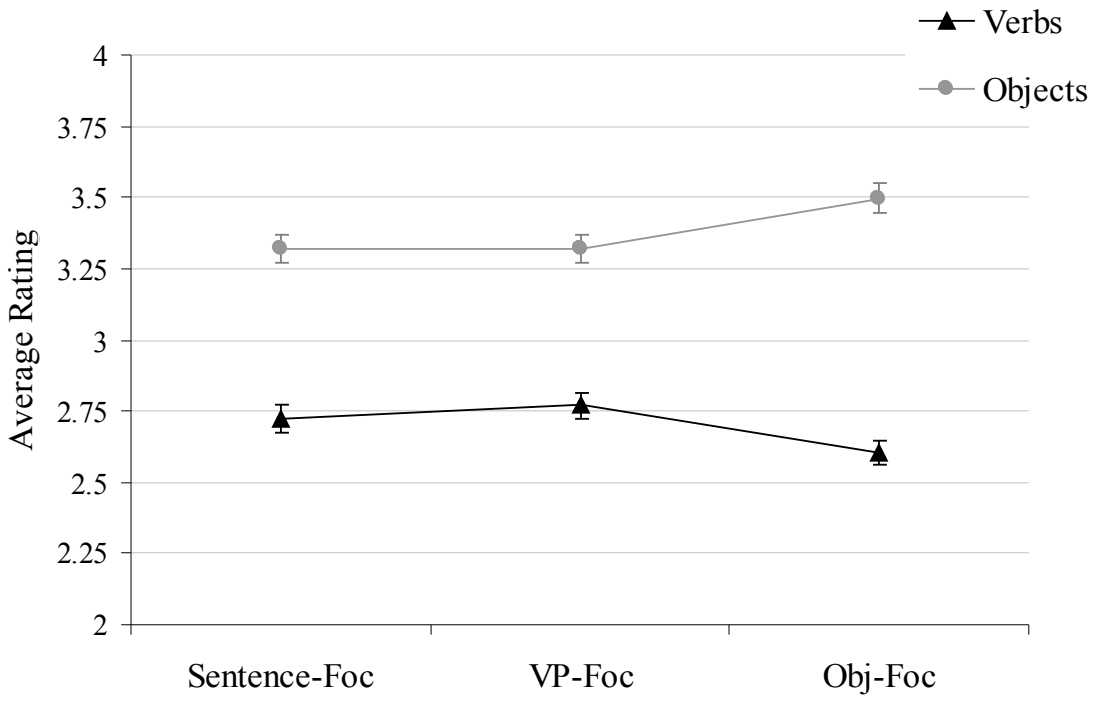

Figure 4: Average prominence ratings for verbs and objects in test sentences in the three focus conditions in Experiment 2. '1' is lowest in prominence, '5' is highest. Error bars show standard error.

4.2.3. Discussion

In relation to focus size, the pattern of results in Experiment 2 did not differ in any significant way from those in Experiment 1; narrowly focused objects were perceived as more prominent than objects in broader focus constituents when focus was contrastive, and there was no clear indication that this effect was more (or less) robust than was the case for non-contrastive focus. Also as in Experiment 1, the boosting in prominence ratings for narrowly focused objects accompanied a suppression of prominence ratings for prenuclear verbs. A final point to make here is that it is, again, unlikely that listeners were relying on expectations alone. In Experiment 2 there were a greater proportion of stimuli that lacked any accentuation on the prenuclear verb in the test stimuli (as indicated by the ToBI labels assigned by transcribers). If listeners were attending to the signal, we would expect their ratings to be influenced by this fact about the signal presented to them. In fact, the significant main effect for *word* indicates this was the case. Further, the difference was much larger between verb and object ratings in Experiment 2 compared with Experiment 1, which is consistent with the characteristics of the stimuli presented to listeners in the two experiments. This indicates that listeners' ratings were highly correlated with aspects of the signal, and their information structural expectations about the signal were an additional influence.

5. General discussion

In the two experiments presented above, English-speaking listeners' expectations about prominence patterns in SVO constructions were shown to be systematically different depending on their interpretation of their information structure. For both non-contrastive and explicitly contrastive focus types, a sentence-final object was judged as more prominent when under narrow focus than when part of a broader sentence or VP focus constituent. What we should regard as a crucial piece of evidence is a particular detail about listeners' expectations, namely that prominence was relative; expectations were not simply for a more prominent nuclear accent on the object, but also a less prominent verb. Our prediction throughout this study was that precisely this pattern should result if listeners have some kind of conventionalized knowledge about how speakers are known to produce foci of different sizes. It might additionally be pointed out that where we did *not* find differences is also informative. That is, production studies that have tested both VP and sentence focus have generally not reported speakers to distinguish the two prosodically, and listeners in the present study correspondingly did not seem to expect them to. Finally, it should be reiterated that listeners' expectations about prominence modulated – but did not completely "override" – their perception of the signal; this can be seen in the relative ratings for verbs and objects in each experiment (i.e., the main effect for "word"), and the size of this difference across the two experiments.

There are two important questions that the results of the present study raise. The first is why listeners seem to have such strong expectations (strong enough to have an illusory effect) for a distinction that is reported to be so phonetically subtle on average. It might be noted, however, that listeners' assumption about prototypical representations tend to be hyperarticulated ones. A clear demonstration of this is the "Hyperspace Effect" for vowels. Using a task in many ways similar to the one used in Experiment 2 in Rump and Collier's (1996) study, Johnson, Flemming and Wright (1993) asked American English-speaking participants to manipulate simple parameters on a speech synthesizer to create a series of vowels. These vowels were to sound like those in various English example words (e.g. "heed", "hid", "aid", "had", "who'd", etc). What the authors found was that participants

consistently adjusted their syntheses until each vowel had considerably peripheral F1 and F2 values. Further, those syntheses exaggerated subtle distinctions such as the spectral difference between [eɪ] and [ɪ]. Yet while participants' synthetic vowel spaces were more extreme and dispersed than their normal citation-form productions (the same participants also produced the vowels), they were in fact within the range of those produced in hyperarticulated contexts. Johnson and colleagues suggested this finding indicated that hyperarticulated pronunciations are the ones used by listeners to internally represent the linguistic distinction between vowel categories. Thus it is possible that listeners' behavior in our current experiments similarly reflected their representations of broad and narrow focus sentences.

This issue feeds into a second, however, i.e., why studies have often failed to find any differences at all for broad versus narrow focus (in terms of both production and perception). This matter may be more methodologically complicated than traditionally thought. Given how subtle the differences often are, and the kinds of variation they are sensitive to, it may be that speakers do not produce them reliably or as saliently when the context is available (Lieberman 1963, 1967; Snedeker and Trueswell 2003). It is also possible (although somewhat less likely; see Snedeker and Trueswell 2003) that listeners might correspondingly be less sensitive to subtle phonetic cues to focus size if the context is available. Complicating matters even further is recent evidence suggesting that how and whether listeners exploit various levels of linguistic context is subject to individual differences in cognitive processing styles (Yu 2010) and possibly pragmatic skills (Bishop in prep.).

However, at an even more fundamental level, there are reasons to question the subtlety production and perception studies usually report. This is due mainly to the possibility that the traditional method of elicitation, i.e., reading printed materials, is not optimal for the subject matter. It is known that there are performance issues with reading aloud that make it different from, for example, silent reading (Bookheimer et al. 1995; see also Jun 2010), and it seems more than likely that it is also different from speaking for genuinely communicative purposes. This would have obvious implications not only for the production studies discussed in Section 2, but also the perception studies that relied on read materials for stimuli. Evidence that this should be regarded as a real concern comes from the fact that the largest (and, to listeners, most useful) differences that have been reported for broad versus narrow focus sentences are in fact from the study that relied least on read materials (Breen et al. 2010). Thus, it is possible that that the standard practice of using such materials, and doing so in uncommunicative situations, will be unsuitable for investigating the prosodic realization of syntactic/information/prosodic structures with a high degree of "surface" ambiguity.

Finally, it is not within the scope of the present study to make sweeping, and especially cross-linguistic, claims about the prosodic representation of focus generally, as we have looked at only one information structural contrast and in only one language. However, what we have considered here suggests that any theory that models the knowledge that English speakers and listeners share in terms of the distribution of pitch accents is, at best, incomplete. Theories that "project" focus to higher syntactic structures depending on the location of the nuclear pitch accent (e.g. Gussenhoven 1984; Selkirk 1995), while successful in capturing a wide range of facts, predict a genuine ambiguity in the SVO structures we have considered here. Consequently, such models seem to capture neither the details of what speakers do, nor, as we have shown here, what listeners know. Thus the present study adds to a growing body of results demonstrating that models of the information structure–prosody relationship will need to consider more detailed aspects of prosodic realization.

6. Conclusion

What prosodic structure a speaker will assign to a sentence is partially dependent on the sentence's information structure, which in turn is dependent on the context in which it is uttered. The present study explored the consequences of this prosody-meaning relationship for the listener. We probed listeners' expectations for knowledge of a relationship between the size of the focus constituent and patterns of prosodic prominence in simple English SVO constructions. We found that listeners responded to broad and narrow focus in a systematically different way. Although they were presented with the same acoustic information, listeners judged an object as significantly more prominent, and a preceding verb as less prominent, when that object was under narrow focus. Here it was suggested that our answer to why this should be the case can be found in listeners' experience with speakers' productions. It was noted that this very pattern, although subject to much variation, is precisely what we would predict given previous production studies. Indeed, it was suggested that, given the methods previous studies have employed, it is possible that the reported differences might actually underestimate those in the listener's experience, i.e., those produced by speakers in natural and communicative contexts. Taken as a whole, the results of the study indicate that English-speaking listeners know more about how prosody can be used to express the information structure of a sentence than is encoded in the distribution of intonational pitch accents.

Appendix: Stimuli and contexts

1.) ***(No… because) I bought a motorcycle.***

	Non-Contrastive Context	Contrastive Context
S-Foc	What happened yesterday?	Why's your wife mad? Because your roof's leaking?
VP-Foc	What did you do yesterday?	Why's your wife mad? Because you lost your job?
Obj-Foc	What did you buy yesterday?	Why's your wife mad? Because you bought a car?

2.) ***(No… because) I lost my wallet.***

	Non-Contrastive Context	Contrastive Context
S-Foc	What happened?	Why are you upset? Because of the economy?
VP-Foc	What did you do?	Why are you upset? Because you overslept?
Obj-Foc	What did you lose?	Why are you upset? Because you lost your keys?

3.) ***(No… because) I failed my midterm.***

	Non-Contrastive Context	Contrastive Context
S-Foc	What happened?	Why are you so worried? Because of the GREs?
VP-Foc	What did you do?	Why are you so upset? Because you're running late?
Obj-Foc	Your grade is really low... what did you fail?	Why are you so upset? Because you failed your quiz?

4.) ***(No… because) I met a girl.***

	Non-Contrastive Context	Contrastive Context
S-Foc	What happened?	Why are you so happy? Because school's out?
VP-Foc	What did you do at the party last night?	Why's your mom so excited? Because you graduated?
Obj-Foc	Who did you meet at the party last night?	Why are you so happy? Because you met a movie star?

5.) ***(No… because) I read a book.***

	Non-Contrastive Context	Contrastive Context
S-Foc	What happened at home?	Why were you up so late? Because it was noisy?
VP-Foc	What did you do at home?	Why were you up so late? Because you were doing homework?
Obj-Foc	What did you read at home?	Why were you up so late? Because you read a magazine?

6.) ***(No… because) I passed the final.***

	Non-Contrastive Context	Contrastive Context
S-Foc	What happened in class?	Why are you so happy? Because it's Friday?
VP-Foc	What did you do in class?	Why are you so happy? Because you finished reading?
Obj-Foc	What did you pass?	Why are you so happy? Because you passed the quiz?

7.) ***(No… because) I bought a car.***

	Non-Contrastive Context	Contrastive Context
S-Foc	What happened while I was gone?	Why are you so broke all of the sudden? Because of the economy?
VP-Foc	What did you do with all your money?	Why are you so broke? Because you started gambling?
Obj-Foc	What did you buy with all your money?	Why are you so broke? Because you bought a house?

8.) ***(No… because) I rode a Harley.***

	Non-Contrastive Context	Contrastive Context
S-Foc	What happened today?	Why are you so excited? Because of the game?
VP-Foc	What did you do?	Why are you so excited? Because you went jogging?
Obj-Foc	What did you ride?	Why are you so excited? Because you rode a pony?

9.) ***(No… because) I bought a watch.***

	Non-Contrastive Context	Contrastive Context
S-Foc	What's up?	Why are you so happy? Because of the weather?
VP-Foc	What did you do?	Why are you so happy? Because you talked to Suzie?
Obj-Foc	What did you buy?	Why are you so happy? Because you bought a hat?

10.) ***(No… because) I drove a Porsche.***

	Non-Contrastive Context	Contrastive Context
S-Foc	What happened?	Why are you so happy? Because of the party today?
VP-Foc	What did you do?	Why are you so happy? Because you went shopping?
Obj-Foc	What did you drive?	Why are you so happy? Because you drove a Mercedes?

11.) ***(No… because) I finished my paper.***

	Non-Contrastive Context	Contrastive Context
S-Foc	What happened?	Why are you so happy? Because class was cancelled?
VP-Foc	What did you do?	Why are you so happy? Because you went on a date?
Obj-Foc	What did you finish?	Why are you so happy? Because you finished your homework?

12.) ***(No… because) I ate a hamburger.***

	Non-Contrastive Context	Contrastive Context
S-Foc	What happened at the picnic?	Why aren't you hungry? Because of the medication?
VP-Foc	What did you do at the picnic?	Why aren't you coming to lunch? Because you're dieting?
Obj-Foc	What did you eat at the picnic?	Why aren't you hungry? Because you ate a hot dog?

13.) ***(No… because) I called the doctor.***

	Non-Contrastive Context	Contrastive Context
S-Foc	What happened?	Why are you feeling so much better? Because of the weather?
VP-Foc	What did you do?	Why are you feeling so much better? Because you slept in?
Obj-Foc	Who did you call?	Why are you feeling so much better? Because you called the nurse?

14.) ***(No… because) I pawned the stereo.***

	Non-Contrastive Context	Contrastive Context
S-Foc	What happened?	Why are you so rich all of the sudden? Because of the stimulus check?
VP-Foc	What did you do?	Why are you so rich all of the sudden? Because you worked overtime?
Obj-Foc	What did you pawn?	Why are you so rich all of the sudden? Because you pawned the T.V.?

15.) ***(No… because) I fixed the roof.***

	Non-Contrastive Context	Contrastive Context
S-Foc	What happened?	Why's your wife so happy? Because of the vacation?
VP-Foc	What did you do?	Why's your wife so happy? Because you took her out to dinner?
Obj-Foc	What did you fix?	Why's your wife so happy? Because you fixed the fence?

16.) ***(No… because) I painted the kitchen.***

	Non-Contrastive Context	Contrastive Context
S-Foc	What happened?	Why's your wife so happy? Because it's her birthday?
VP-Foc	What did you do?	Why were you busy all day? Because you were working out?
Obj-Foc	What did you paint?	Why's your wife so happy? Because you painted the fence?

17.) ***(No… because) I kissed another cheerleader.***

	Non-Contrastive Context	Contrastive Context
S-Foc	What happened after the game?	Why are you smiling like that? Because of the game?
VP-Foc	What did you do after the game?	Why are you smiling like that? Because you played well?
Obj-Foc	Who did you kiss after the game?	Why are you smiling like that? Because you kissed another pompom girl?

Notes

[1] This term is preferred to another common one, the "domain of focus", for primarily two reasons. First, it is a slightly more theoretically neutral term. Second, we wish to distinguish it from another common use of "domain" in the literature on focus, which refers to the syntactic or semantic domain of a focus sensitive operator such as "only".

[2] This distinction is made with the understanding that there is a long-standing and still unsettled debate as to whether contrast versus non-contrast is a grammatical distinction (in English). While some authors have assumed the grammar encodes the difference (Chafe 1976; Vallduví and Vilkuna 1998, among others), others have suggested a single category, all focus being essentially a kind of contrast (e.g. Jackendoff 1972; Rooth 1992). A slightly different view comes from Büring (2007), who suggests that the distinction is not one of grammar, but of usage – a matter of interpreting a speakers' intentions in a particular pragmatic context. Although the results of the experiments presented in Section 4 may be taken as relevant to the debate, a contribution to it is not a primary goal of the present paper. (For a recent discussion of the matter, however, see Katz and Selkirk, submitted).

[3] Some of the phonetic differences we discuss for focus size have not been reported, for example, for some Romance languages (e.g. D'Imperio 1997; Frota 2000, 2002). However, Jun (2008) has presented production evidence from Korean that suggests the basic patterns reported for English, Dutch and German may have strikingly close analogues even in languages with very different prominence-marking systems. Therefore, although a larger sample of languages is needed, it does not seem the phenomenon that we explore in the present paper is confined to West Germanic.

[4] A similarly independent effect for the number of repetitions was not yet explored by the authors.

[5] The ToBI system for transcribing intonation and prosody (Beckman and Hirschberg 1994), like a number of other such systems, including RaP (Dilley and Brown 2005) and ToDI (Gussenhoven 2005), is based on a phonological model of the target language. As such, it necessarily assumes an indirect acoustics-meaning relationship, which is not uncontroversial (e.g. Xu and Xu 2005). Nothing presented here crucially depends on this matter, however, and for the present purposes use of the transcriptions is mostly practical; MAE_ToBI is a widely used standard for prosodic transcription of American English, and it is assumed an effective way to communicate the form of the stimuli used in the present study.

[6] In one case, the second labeler annotated an object as ambiguous between a !H* and a delayed L*. The !H* annotation was used here to calculate agreement between raters.

Acknowledgements

This paper has greatly benefited from discussions with Sun-Ah Jun, Daniel Büring, Carlos Gussenhoven, Patricia Keating, Robert Daland, Jennifer Zhang, members of the UCLA Phonetics Laboratory, and especially from the thoughtful comments and suggestions provided by two anonymous reviewers. The author would also like to thank Natasha Abner, Craig Sailor, Diana Hill and Mariam Bassali for help with materials and data collection, and the UCLA ATS Statistics Consulting Group for their advice on statistical modeling. Finally, I am grateful to Gorka Elordieta, Pilar Prieto and organizers at the Universitat Pompeu Fabra, Universitat Autònoma de Barcelona and the Institut d'Estudis Catalans for hosting the lively and very stimulating Workshop on Prosody and Meaning in Barcelona in 2009.

References

Bartels, Christine and John Kingston (1994). Salient pitch cues in the perception of contrastive focus. In: Peter Bosch and Rob van der Sandt (eds.), *Focus and natural language processing, Volume 1: Intonation and Syntax*, 1-10. IBM Deutschland Informationssysteme GmbH Scientific Center, Institute for Logic and Linguistics.

Baumann, Stefan, Johannes Becker, Martine Grice and Doris Mücke (2008). Tonal and articulatory marking of focus in German. *Proceedings of the XVIth International Congress of Phonetic Sciences*, 1029-1032. Saarbrücken, Germany.

Baumann, Stefan, Martine Grice and Susanne Steindamm (2006). Prosodic marking of focus domains: Categorical or gradient? *Proceedings of Speech Prosody 2006*, 301-304. Dresden, Germany.

Beckman, Mary (1986). *Stress and Non-stress Accent.* Netherlands Phonetic Archives 7. Dordrecht: Foris.

Beckman, Mary and Julia Hirschberg (1994). The ToBI annotation conventions. Ms. The Ohio State University.

Beckman, Mary (1996). The parsing of prosody. *Language and Cognitive Processes* 11: 17-67.

Birch, Stacy and Charles Clifton (1995). Focus, accent and argument structure: Effects on language comprehension. *Language and Speech* 38: 365-391.

Bishop, Jason (in prep). Information structural interpretation in on-line sentence processing: focus, prosody, and individual differences in listeners' "autistic" traits. Ms. University of California, Los Angeles.

Bookheimer, Susan, Thomas Zeffiro, Teresa Blaxton, William Gaillard and William Theodore (1995). Regional cerebral blood flow during object naming and word reading. *Human Brain Mapping* 3: 93-106.

Breen, Mara, Evelina Fedorenko, Michael Wagner, and Edward Gibson (2010). Acoustic correlates of information structure. *Language and Cognitive Processes* 25: 1044-1098.

Breen, Mara, Laura Dilley, John Kraemer and Edward Gibson (in press). Inter-transcriber agreement for two systems of prosodic annotation: ToBI (Tones and Break Indices) and RaP (Rhythm and Pitch). *Corpus Linguistics and Linguistic Theory.*

Büring, Daniel (2007). Intonation, semantics and information structure. In: Gillian Ramchand and Charles Reiss (eds.), *The Oxford Handbook of Linguistic Interfaces*, 445-473. Oxford: Oxford University Press.

Calhoun, Sasha (2006). Information structure and the prosodic structure of English: A probabilistic relationship. PhD. Dissertation, University of Edinburgh.

Chafe, Wallace (1976). Givenness, contrastiveness, subject, topic, and point of view. In Charles N. Li (ed.), *Subject and Topic*, 25-55. New York: Academic Press.

Cole, Jennifer, Yoonsook Mo and Mark Hasegawa-Johnson (2010). Signal-based and expectation-based factors in the perception of prosodic prominence. *Laboratory Phonology* 1: 425-452.

Dilley, Laura and Meredith Brown (2005). The RaP (Rhythm and Pitch) Labeling System, Version 1.0. Available at http://tedlab.mit.edu/rap.html.

D'Imperio, Mariapaola (1997). Breadth of focus, modality and prominence perception in Neapolitan Italian. *Ohio State University Working Papers in Linguistics* 50: 19-39.

Dupoux, Emmanuel, Kazohiko Kaheki, Yuki Hirose, Christophe Pallier and Jacques Mehler (1999). Epenthetic vowels in Japanese: A perceptual illusion? *Journal of Experimental Psychology: Human Perception and Performance* 25: 1568-1578.

Eady, Stephen, William Cooper, Gayle Klouda, Pamela Mueller, Dan Lotts (1986). Acoustical characteristics of sentential focus: Narrow vs. broad and single vs. dual focus environments. *Language and Speech* 29: 233-251.

Eriksson, Anders, Gunilla Thunberg, and Hartmut Traunmüller (2001). Syllable prominence: A matter of vocal effort, phonetic distinctness and top-down processing. *Proceedings of the European Conference on Speech Communication and Technology*, 399-402. Aalborg, Denmark.

Frota, Sónia (2000). *Prosody and Focus in European Portuguese: Phonological Phrasing and Intonation.* New York: Garland Publishing.

Frota, Sónia (2002). The prosody of focus: a case-study with cross-linguistic implications. *Proceedings of Speech Prosody 2002*, 319-322. Aix-en-Provence, France.

Gussenhoven, Carlos (1983). Testing the reality of focus domains. *Language and Speech* 26: 61-80.

Gussenhoven, Carlos (1984). *On the Grammar and Semantics of Sentence Accents*. Dordrecht: Foris.

Gussenhoven, Carlos, B. Repp, A. Rietveld, H. Rump and J. Terken (1997). The perceptual prominence of fundamental frequency peaks. *Journal of the Acoustical Society of America* 102: 3009-3022.

Gussenhoven, Carlos (2005). Transcription of Dutch Intonation. In: Jun, Sun-Ah (ed.), *Prosodic Typology: The Phonology of Intonation and Phrasing*, 118-145. Oxford: Oxford University Press.

Gussenhoven, Carlos (2008). Notions and subnotions in information structure. *Acta Linguistica Hungarica* 55: 381-395.

Hanssen, Judith, Jörg Peters and Carlos Gussenhoven (2008). Prosodic effects of focus in Dutch declaratives. *Proceedings of Speech Prosody 2008*, 609-612. Campiñas, Brazil.

He, Xuliang, Judith Hanssen, Vincent van Heuven and Carlos Gussenhoven (2011). Phonetic implementation must be learnt: Native versus Chinese realization of focus accent in Dutch. *Proceedings of the XVIIth International Congress of Phonetic Sciences*, 843-846. Hong Kong.

Ito, Kiwako, Shari Speer and Mary Beckman (2004). Informational status and pitch accent distribution in spontaneous dialogues in English. *Proceedings of Speech Prosody 2004*, 279-282. Nara, Japan.

Ito, Kiwako and Shari Speer (2008). Anticipatory effects of intonation: Eye movements during instructed visual search. *Journal of Memory and Language* 58: 541-573.

Jackendoff, Ray (1972). *Semantics in Generative Grammar*. Cambridge, MA: MIT Press.

Johnson, Keith, Edward Flemming and Richard Wright (1993). The hyperspace effect: Phonetic targets are hyperarticulated. *Language* 69: 505-527.

Jun, Sun-Ah (2008). Focus: domains, types, and realizations. Talk given at the Yale Linguistics Department Colloquium Series, Yale University, New Haven, CT.

Jun, Sun-Ah (2010). The implicit prosody hypothesis and over prosody in English. *Language and Cognitive Processes* 25: 1201-1233.

Katz, Jonah and Elisabeth Selkirk (submitted). Contrastive focus vs. discourse-new: Evidence from prosodic prominence in English. Ms. Department of Linguistics, MIT.

Kochanski, Greg, Ester Grabe, John Coleman and Burton Rosner (2005). Loudness predicts prominence: Fundamental frequency lends little. *Journal of the Acoustical Society of America* 118, 1038-1054.

Krifka, Manfred (2008). Basic notions of information structure. *Acta Linguistica Hungarica* 55(3-4): 243-276.

Ladd, Robert (1980). *The Structure of Intonational Meaning: Evidence from English*. Bloomington, IN: Indiana University Press.

Ladd, Robert (1996). *Intonational Phonology*. Cambridge: Cambridge University Press.

Lieberman, Philip (1963). Some effects of semantic and grammatical context on the production and perception of speech. *Language and Speech* 6: 172-187.

Lieberman, Philip (1967). *Intonation, Perception and Language*. Cambridge, MA: MIT Press.

Löfstedt, Ingvar (2006). On focus types and focus domains. Ms. University of California, Los Angeles.

Mo, Yoonsook (2008). Duration and intensity as perceptual cues for naïve listeners' prominence and boundary perception. *Proceedings of Speech Prosody 2008*, 639-742. Campiñas, Brazil.

Pitt, Mark, Laura Dilley, Keith Johnson, Scott Kiesling, William Raymond, Elizabeth Hume and Eric Fosler-Lussier (2007). *Buckeye Corpus of Conversational Speech* (2nd release) [www.buckeyecorpus.osu.edu] Columbus, OH: Department of Psychology, Ohio State University (Distributor).

Pierrehumbert, Janet (1979). The perception of fundamental frequency declination. *Journal of the Acoustical Society of America* 66: 363-369.

Rietveld, Toni and Carlos Gussenhoven (1985). On the relation between pitch excursion size and pitch prominence. *Journal of Phonetics* 13: 299-308.

Rump, Hans and René Collier (1996). Focus conditions and the prominence of pitch-accented syllables. *Language and Speech* 39: 1-17.

Rooth, Mats (1992). A theory of focus interpretation. *Natural Language Semantics* 1, 75–116.

Samuel, Arthur (1981). Phonemic restoration: insights from a new methodology. *Journal of Experimental Psychology: General* 110: 474-494.

Selkirk, Elisabeth (1984). *Phonology and Syntax: The Relation between Sound and Structure*. Cambridge MA: MIT Press.

Selkirk, Elisabeth (1995). Sentence prosody: Intonation, stress, and phrasing. In: John A. Goldsmith (ed.), *The Handbook of Phonological Theory*, 550-569. Oxford: Blackwell.

Sityaev, Dmitry and Jill House (2003). Phonetic and phonological correlates of broad, narrow and contrastive focus in English. *Proceedings of the XVth International Congress of Phonetic Sciences,* 1819-1822. Barcelona.

Snedeker, Jesse and John Trueswell (2003). Using prosody to avoid ambiguity: Effects of speaker awareness and referential context. *Journal of Memory and Language* 48: 103-130.

Syrdal, Ann and Julia McGory (2000). Inter-transcriber reliability of ToBI prosodic labeling. *Proceeding of ICSLP 2000*, 235-238. Beijing.

Vallduví, Enric and Maria Vilkuna (1998). On rheme and kontrast. *Syntax and Semantics* 29: 79-107.

Welby, Pauline (2003). Effects of pitch accent position, type, and status on focus projection. *Language and Speech* 46: 53-81.

Warren, Richard (1970). Perceptual restoration of missing speech sounds. *Science* 167: 392-393.

Xu, Yi and Ching Xu (2005). Phonetic realization of focus in English declarative intonation. *Journal of Phonetics* 33: 159-197.

Yu, Alan (2010). Perceptual compensation is correlated with individuals' "autistic" traits: Implications for models of sound change. *PLoS ONE* 5(8) [e11950.doi:10.1371/journal.pone.0011950].