



3-1-2015

Prosody Marks Different Kinds of Informativity: Interactions Between Frequency, Probability and Focus

Iris Ouyang
iris.oy@gmail.com

Elsi Kaiser

Prosody Marks Different Kinds of Informativity: Interactions Between Frequency, Probability and Focus

Prosody marks different kinds of informativity: Interactions between frequency, probability and focus

Iris Chuoying Ouyang and Elsi Kaiser

1 Introduction

It is widely accepted that the prosody of a sentence can signal the extent to which a word in the sentence is ‘informative’. The acoustic-phonetic properties of an utterance such as duration, pitch/f₀, intensity, and spectral characteristics provide cues for the relative informativity of its components (see Wagner & Watson 2010 for a review). A lot of prior work has approached the prosodic encoding of informativity either from (i) the vantage point of information structure, within a discourse-pragmatic approach (e.g., Breen, Fedorenko, Wagner and Gibson 2010, Chen and Brown 2006, Cooper, Eady and Mueller 1985, Couper-Kuhlen 1984, Eady, Cooper, Klouda, Mueller and Lotts 1986, Katz and Selkirk 2011, Krahmer and Swerts 2001, Schwarzschild, 1999) or (ii) from the vantage point of information theory, with a focus on statistical probability (e.g., Aylett and Turk 2004, Baker and Bradlow 2009, Bell, Brenier, Gregory, Girand and Jurafsky 2009, Bell, Jurafsky, Fosler-Lussier, Girand, Gregory & Gildea 2003, Munson & Solomon 2004, Pan and Hirschberg 2000, Pluymaekers, Ernestus & Baayen 2005, van Son, Koopmans-van Beinum and Pols 1998, Wright 2003).

In the **information-structure-based tradition**, acoustic prominence is associated with linguistic material in the foreground, or in focus—broadly speaking, material that adds new information to the discourse. Various acoustic properties reflect information-structural salience, including pitch accents (e.g., Pierrehumbert and Hirschberg 1990), increased duration, pitch and intensity, and larger excursions of pitch and intensity (Breen et al. 2010, Katz and Selkirk 2011, Chinese Mandarin: Ouyang and Kaiser 2013). Elements in *new-information focus* (new to the discourse) are prosodically more prominent than given/old elements (already-mentioned information) (e.g., Chen and Braun 2006, Schwarzschild 1999). More specifically, words in *narrow* new-information focus—where only one part of the sentence (e.g., action, patient, location) is focused—receive greater prosodic prominence than words in *wide* new-information focus, where the entire utterance is in focus (e.g., Breen et al. 2010, Eady et al. 1986). It has also been shown that elements in contrastive focus (i.e. which have alternatives in some way), are realized with greater prosodic prominence than non-contrastive elements (e.g., Cooper et al. 1985, Couper-Kuhlen 1984, Katz and Selkirk 2011, Krahmer and Swerts 2001). Existing cross-linguistic work has identified various types of contrastive focus (e.g., Vallduví and Vilkuna 1998). We concentrate on one subtype, *corrective focus*, which corrects previously-conveyed information (see Breen et al. 2010, Dik 1997). We chose this type of contrastive focus because its information-structural properties are well-understood and it is prevalent in communication.

On the other hand, research in the **information theory-based tradition** has identified a correlation between acoustic reduction and the redundancy or the predictability of linguistic material. Elements with higher predictability are acoustically more reduced, including being realized with decreased duration (e.g., Bell et al. 2003), no pitch accent (e.g., Pan and Hirschber 2000), lower center of gravity of the power spectrum (CoG), less extreme distance between the first and second formants (e.g., van Son et al. 1999) and less dispersed vowel space (e.g., Wright 2003, Munson and Solomon 2004). Probabilistic measurements used to represent the predictability of a linguistic element include context-independent properties such as lexical frequency and neighborhood density (e.g., Munson and Solomon 2004, Wright 2003) and context-dependent properties such as joint probability and conditional probability (e.g., Bell et al. 2003, Pan and Hirschberg 2000, van Son et al. 1998).

Are there connections between information-structural and information-theoretic approaches? These two traditions have investigated rather different factors of informativity from distinct perspectives, despite a common interest in informativity and prosody. A limited number of studies in the information-theoretic approach have examined consequences of repeated mention (e.g., Aylett and Turk 2004, Baker and Bradlow 2009, Bell et al. 2009, Pluymaekers et al. 2005). Repeated words are by definition given information, and thus the information-theoretic notion of

repetition can be regarded as givenness in an information-structural view. As far as we know, only one prior study has addressed the interaction between information-theoretic properties and information structure: Baker and Bradlow (2009) found that word frequency influences reduction in repeated words; high-frequency words exhibit more shortening upon second mention than low-frequency words. It remains unclear whether other kinds of information-theoretic factors also have an impact on the prosodic effects of information structure, and whether word frequency interacts with other types of information structure in a similar way. In essence, it is not yet well understood whether information-theoretic factors and information structure interact to influence prosody.

1.1 Aims of Our Study

As mentioned in the Introduction, existing research has found that an utterance's prosodic representation depends on how informative the constituents are relative to each other. Information-structural properties, such as discourse focus, and information-theoretic properties, such as lexical frequency and contextual probability, both play a role in how informativity is encoded prosodically. It is striking that little attention has been paid to the potential interaction between information structure and information-theoretic factors, given the considerable efforts that have been devoted to both kinds of factors separately. To shed light on this issue, we conducted a psycholinguistic production study with three inter-related aims: (i) Do information structure and information-theoretic factors interact in determining a word's prosodic prominence, and if so, how? (ii) If information-theoretic factors do interact with information structure in shaping the prosody of a sentence, to what extent can information-theoretic factors impact the prosodic effects of information structure? (iii) If the prosodic encoding of information structure depends on a word's information-theoretic properties, do different information-structural categories interact with different information-theoretic factors in similar ways? For instance, could it be that the prosodic cues for new-information vs. corrective focus differ in terms of whether they are sensitive to lexical frequency vs. contextual probability?

Theoretically, there are three hypotheses that can be formulated with regard to whether and how information-theoretic factors impact the prosodic effects of information structure in general: First, it could be that information-theoretic factors and information structure do not interact (Hypothesis I). More specifically, it could be that different information-structural categories are always marked distinctively in prosody, and these prosodic differences remain constant regardless of the information-theoretic properties.

However, if information-theoretic factors and information structure do interact, there are at least two ways in which this could occur. It could be that the information-structural properties of a word influence its prosodic prominence, but this effect is weak in words with low informativity in the information-theoretic dimensions (Hypothesis II). For example, narrow and wide focus might greatly differ in prosody when the focused is a low-frequency word (i.e., has high information-theoretic informativity) but not so much when the focused word is a very frequent word (i.e., has low information-theoretic informativity).

Alternatively, it could be that this effect of information structure is weak in words with high informativity in the information-theoretic dimensions (Hypothesis III). In other words, it could be that prosodic effects of information structure are weakened when there are other factors—such as information-theoretical measures of informativity—demanding prosodic prominence. For example, the prosodic cues for narrow and wide focus might be clearly distinct when the focused word is highly frequent (i.e., has low information-theoretic informativity) but not so much when the focused word is infrequent (i.e., has high information-theoretic informativity).

Since existing work has shown a stronger effect of givenness on duration when the repeated words are high in frequency (Baker and Bradlow 2009), we believe Hypothesis III to be the most likely: While different information-structural categories can be distinguished by prosody, these distinctions may be masked in words that are highly informative in some information-theoretic aspect. Building on Baker and Bradlow (2009), our study explored effects of **word frequency** and **narrow new-information focus**. We also looked at the effects of another information-theoretic factor—**contextual probability**—as well as another information-structural category—**narrow corrective focus**. We compared the two types of narrow focus against **wide/VP focus** (see details in the Method section). With two information-theoretic factors and three information-structural

categories included in this study, we examined whether the effects of information structure varied depending on the information-theoretic condition and whether different information-theoretic factors (word frequency and contextual probability) impacted the effects of different information-structural categories (new-information focus and corrective focus) in different ways.

This paper builds on Ouyang and Kaiser (2014), which describes an earlier version of this research. In our earlier work, we explored the notion of perspective-taking and the question of whether speakers take into account the interlocutor's internal knowledge state. In the current paper, we introduce the notion of 'epistemic surprisal,' which we think offers a more accurate way of capturing our results, while minimizing the need for additional assumptions. We use the term 'epistemic surprisal' to refer to the extent to which the speaker's expectations diverge from what actually happens. A higher degree of epistemic surprisal means a bigger mismatch between what has just happened/what has been said and what the speaker had expected. As will become clear in the Discussion section, we show how our results—in particular, the prosodic patterns produced by speakers—can be explained as a result of speakers' expectations and surprise about their conversational partner's utterances. We will show that the interplay of lexical frequency, contextual probability and information structure can be regarded as an effect of 'epistemic surprisal'.

2 Method

We conducted a production study with an interactive set-up. Each trial consisted of a production task and a subsequent selection task. In both tasks, participants interacted with a partner (who was a lab assistant). Sixteen native speakers of American English participated. The production task provided the critical recordings: The target sentences were produced by the participants during the production task. The selection task was included in order to engage both people in the production task: Paying attention to what the other person said in the production task was necessary to successfully perform the selection task. (We do not discuss the selection task in detail due to space limitations and because it is not crucial to the results, but people essentially had to select, with a mouse-click, the correct items from a larger array).

2.1 Production Task

Participants worked with a partner in producing sentence pairs. Each sentence pair consisted of a question asked by the partner (Sentence A) and a response given by the participant (Sentence B), as exemplified in (1-3) below. Participants saw Sentence B on a computer screen when it was their turn to speak. The target sentences (i.e., Sentence B on target trials) are transitive clauses with the following structure: a third-person plural pronoun subject (*they*), a simple past tense verb (e.g., *found*), an object, and a prepositional phrase indicating a location. The critical word we focus on is the object of each target sentence (e.g., *fish*). The experiment had 48 target trials; each participant encountered four items in each condition. Forty-eight filler trials were also included. The dependent variable that we measured was the pitch/f₀ of a sentence.

- (1) NARROW CORRECTIVE FOCUS
A: I heard that Bonnie and Laura found boats in the sea.
B: No, they found [fish]_{CORR FOCUS} in the sea.
- (2) NARROW NEW-INFORMATION FOCUS
A: What did Mary and Irene find in the sea?
B: They found [fish]_{NEW FOCUS} in the sea.
- (3) WIDE/VP FOCUS
A: What did Lillian and Gladys do?
B: They [found fish in the sea]_{NEW FOCUS}.

To investigate whether information-theoretic factors interact with information structure in shaping the prosody of a sentence, we manipulated (i) the lexical frequency of the object, (ii) whether the object was probable in the context of the preceding verb and the following location, and (iii) the object's information structural status, relative to the question. Thus, a within-subject design with three independent variables was implemented: (i) word frequency (with two levels:

high or low frequency), (ii) contextual probability (with two levels: high or low probability), and (iii) information structure (with three levels: narrow corrective focus, narrow new-information focus, or wide/VP focus).

We manipulated the *information structure* of the critical nouns by means of the question asked by the partner (answered by the participant), as exemplified in (1)-(3). In the wide focus (VP focus) condition (3), the question asks about the content of the entire verb phrase (i.e., *what did X do?*) and the answer (spoken by the participant) provides this information. Thus, the whole VP (e.g., *found fish in the sea*) is new information. In the narrow focus new information condition (2), the question asks for the object of the transitive verb and the answer provides this information (e.g., *fish*). Thus, only the object is new information in this condition. Finally, in the narrow focus corrective condition (1), the partner makes a statement where the object is incorrect (as signaled to the participant by the sentence on their screen), and thus the object in the participant's response is correctively focused.

The *word frequency* of the object nouns was determined according to the SUBTLEXus database (Brysbaert and New 2009). The 'incorrect' objects in the questions that elicited corrective focus (1) had word frequencies between the high-frequency and low-frequency critical words for a given context.

2.1.1 Norming Study

The contextual probability of the critical words was estimated through a web-based norming study (conducted on Ibex Farm, <http://spellout.net/ibexfarm/>). Native speakers of American English (who did not participate in the main experiment) performed a fill-in-the-blank writing task: They saw sentences composed of a personal name, a verb, a blank, and a location (e.g., *Christian found _____ in the sea.*) and filled in the blank with one or two words. There were 63 items in total; 66 to 70 responses were collected for each item.

Four verb-location contexts and eight objects were ultimately selected for the target sentences in the main study, as shown in Table 1. For each verb-location context, we identified the most frequently produced nouns, and selected two of the three most frequently produced nouns to act as the contextually high-probability nouns for that context (e.g., *found {fish/shells} in the sea*). In addition, these same nouns were used as low-probability target nouns in other contexts. We made sure that objects used in the low-probability conditions in the main study never occurred as responses for a given context in the norming study (e.g., no one completed *kicked _____ in the garage* with 'books' or 'shells'). In other words, the eight target nouns functioned as contextually high-probability nouns in some contexts and contextually low-probability nouns in other contexts. This allows us to ensure that any effects of contextual probability cannot be attributed to idiosyncratic properties of specific nouns.

Another four nouns were selected to be the 'incorrect' objects mentioned by speaker A in the corrective focus contexts (1). These nouns were responses that did not frequently appear in the norming study but did occur at least once in a given context (e.g., *found boats in the sea*) and thus had a contextual probability between the high-probability and low-probability critical words.

VERB	OBJECT				LOCATION
	Frequent & Probable	Infrequent & Probable	Frequent & Improbable	Infrequent & Improbable	
<i>kicked</i>	<i>cars</i>	<i>cans</i>	<i>books</i>	<i>shells</i>	<i>in the garage</i>
<i>got</i>	<i>balls</i>	<i>cleats</i>	<i>fish</i>	<i>toys</i>	<i>at the sports store</i>
<i>found</i>	<i>fish</i>	<i>shells</i>	<i>balls</i>	<i>cans</i>	<i>in the sea</i>
<i>found</i>	<i>books</i>	<i>toys</i>	<i>cars</i>	<i>cleats</i>	<i>on the stairs</i>

Table 1: Manipulation of word frequency and contextual probability.

2.2 Predictions

Prior work shows that elements in narrow focus are prosodically more prominent than those in wide focus. When it comes to pitch/f₀, this prosodic encoding of narrow focus appears as empha-

sis on the focused element and reduction in the preceding and following elements. In declarative sentences, narrow focus results in higher mean pitch and higher maximum pitch on the stressed syllable of the focused word (e.g., Breen et al. 2010) and lower pitch in the post-focus words (e.g., Cooper et al. 1985) and the pre-focus words (e.g., Japanese: Hwang 2012). Pre-empting our results somewhat, we note here that in our data, the prosodic prominence for narrow focus mostly manifested itself as post-focus reduction, shown by a rapid, substantial drop in pitch after the focused word—a finding in line with the observations of Cooper et al. (1985). Therefore, in the remainder of this paper, we frame our discussion in terms of the pitch movement occurring *after* the object (i.e., the narrowly focused word) in a sentence.

As discussed in Section 1, we hypothesized that *the prosodic effects of information structure may be weakened when there are other factors demanding prosodic prominence*. We want to see whether prosodic cues for information structure would be obscured when information-theoretic informativity was high (low frequency and probability) and whether different information-structural categories (corrective focus and new-information focus) pattern alike.

Based on existing work (e.g., Cooper et al. 1985, Baker and Bradlow 2009), we expected that when the narrowly focused word is highly contextually probable and/or highly frequent, we would be able to observe prosodic reflexes of information structure: Post-focus reduction in the narrow-focus conditions would result in a pitch drop after the object, faster and deeper than the wide-focus condition. However, when the narrowly focused word is low-probability and/or low-frequency, we predicted that the prosodic distinctions between narrow and wide focus might be weakened or even missing: In other words, fast and deep pitch dropping after the object might be observed in only one or perhaps in neither of the two narrow-focus conditions. If these predictions are borne out, we can then look into whether word frequency and contextual probability influence corrective and new-information focus in different ways.

3 Data Analysis and Results

Pitch measurements were obtained in terms of fundamental frequency (f0) (STRAIGHT: Kawahara, de Cheveigne & Patterson 1998, VoiceSauce: Shue, Keating, Vicenik and Yu 2011). The raw f0 values were smoothed (smoothn in MATLAB: Garcia 2010) and converted into a semitone scale using the algorithm $69 + 12 * \log_2(f0/440)$. Semitones have been shown to reflect pitch perception better than the Hertz scale (Nolan 2003). Finally, the data were normalized to z-scores for each participant in order to factor out individual differences in pitch ranges.

The pitch values (i.e., standardized semitones) of target sentences were analyzed using the Smoothing Spline ANOVA method, which performed nonparametric regression on continuous data (Gu 2003). We extracted ten data points with equal time spacing from each of the three intervals in a target sentence: verb (pre-object interval), object, and the region from preposition to article (post-object interval). Mixed-effects models were conducted (*gss* in R: Gu 2013) with Information Structure, Time and their Interaction as fixed effects, and Subjects and Items as random effects. The hypotheses were tested at a significance level of $\alpha=0.05$.

To investigate whether different levels of word frequency and contextual probability influence the prosodic encoding of information structure in different ways, we examined the distinctions between information-structure categories in the four conditions of word frequency and contextual probability separately: high-frequency words in high-probability contexts, high-frequency words in low-probability contexts, low-frequency words in high-probability contexts, and low-frequency words in low-probability contexts.

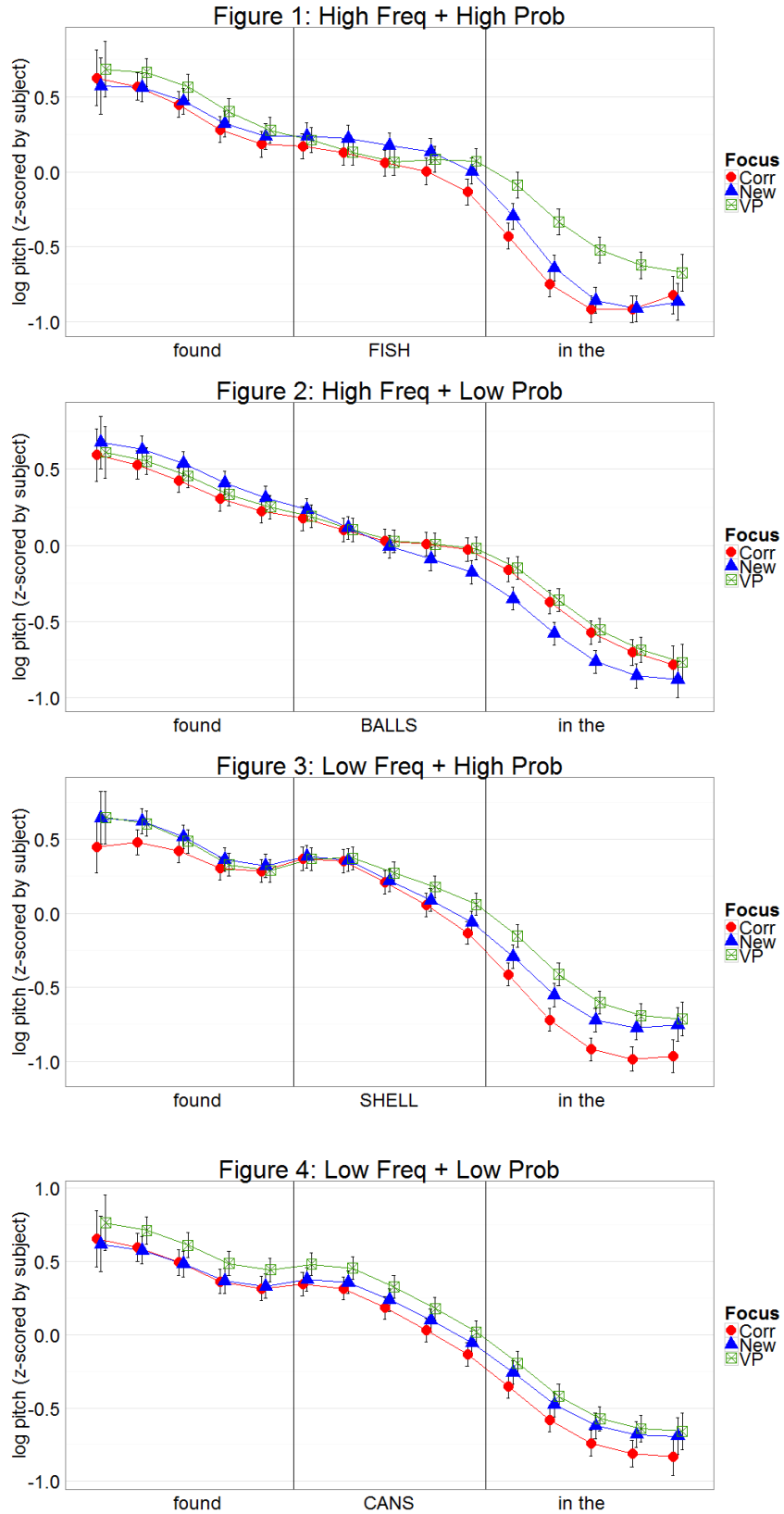


Figure 1-4: Regression modals of log pitch (semitones) in the pre-object, object and post-object regions of a sentence standardized by speaker.

Overall, the predictions outlined in Section 2.2 were borne out, as shown in Figures 1-4, which show fitted values with 1.96 standard errors (95% confidence intervals) obtained from the regression models. The three information-structural categories do not significantly differ in the pre-object and object intervals (first two sections marked on the x-axis). However, significant differences emerge in the post-object interval (last section marked on the x-axis): Narrow corrective focus (red lines with circles) and narrow new-information focus (blue lines with triangles) have a steeper pitch drop than wide focus (green lines with squares) in some cases, depending on the word frequency and contextual probability of the object (the narrowly focused word). More specifically: When the object is a high-frequency word and occurs in a highly probable context (*found fish in the sea*), both types of narrow focus differ significantly from wide focus (Figure 1, labeled ‘High Freq + High Prob’). When the object is a frequent word but contextually improbable, (*found shells in the sea*), only new-information focus differs significantly from wide focus; corrective focus patterns with wide focus (Figure 2, ‘High Freq + Low Prob’). In contrast, when the object has low word frequency but high contextual probability (*found balls in the sea*), corrective focus differs significantly from wide focus; new-information focus does not (Figure 3, ‘Low Freq + High Prob’). Finally, neither type of narrow focus differs from wide focus when the object is infrequent and contextually improbable (*found cans in the sea*, Figure 4, ‘Low Freq + Low Prob’).

4 Discussion

Our experiment investigated how information-theoretic factors and information structure interact in sentence prosody. Prior work has examined prosodic cues from either an information-theoretic or an information-structural perspective, but little work has been done to include these two kinds of informativity factors in one study. Our results show that *the prosodic prominence associated with information structure is modulated by information-theoretic factors*. In particular, the two types of narrow focus are affected by word frequency and contextual probability in different ways. *Corrective narrow focus* results in significant post-focus pitch lowering only when the word carrying corrective information is *probable* in the context. However, *new-information narrow focus* results in significant post-focus pitch lowering only when the word carrying new information is a *frequent* word. When the narrowly focused word is lexically frequent and contextually probable, both types of narrow focus have greater pitch prominence than wide/VP focus. In contrast, when the narrowly focused word is infrequent and improbable, neither narrow focus type is distinguishable from wide/VP focus. This fits with our prediction that prosodic effects of information structure are weakened when other factors demand prosodic prominence.

4.1 Relation to Prior Work

Taken together, our findings challenge the widespread view that narrow focus is (consistently) associated with greater prosodic prominence than wide focus. In fact, existing work on information structure suggests a prominence hierarchy: Contrastive/corrective information is prosodically marked with greater prominence than ‘plain’ new information, and narrowly focused new information is more prominent than new information in wide focus (e.g., Breen et al. 2010, Katz & Selkirk 2011). To the contrary, we did not see this hierarchy in our data. Crucially, it seems that many prior studies have concentrated on relatively probable contexts and have not manipulated word frequency, which may explain the hierarchical relation previously found between corrective focus, new-information focus and wide focus (corrective > narrow new > wide new). Consider a hypothetical study that uses a mix of high-frequency and low-frequency words that are focused in probable contexts. Based on our results, in such a study (a) corrective focus will have greater prominence than wide focus, since the contexts are probable, and (b) new-information focus will be less prominent than corrective focus and more prominent than wide focus, because frequent words pattern with the former but infrequent words with the latter. These predictions are confirmed by a follow-up analysis that we conducted where we pooled the frequency conditions and only included conditions with high contextual probability. Tested at the significance level of $\alpha=0.05$, the pitch fall immediately following the object was largest for corrective focus, second largest for new-information focus, and smallest for wide/VP focus. In other words, the common

generalization that corrective focus is more prominent than narrow focus which is more prominent than wide focus (corrective > narrow new > wide new) may be an epiphenomenon stemming from not controlling word frequency and using relatively probable contexts. Our findings highlight the importance of disentangling information structure and information-theoretic factors: To fully understand how prosody encodes informativity, it is important to integrate the work in the information-theoretic approach and the work in the information-structural approach (see Wagner & Watson 2010:933 for relevant discussion).

4.2 Epistemic Surprisal

In this section, we start to explore the question of *why* information-theoretic factors interact with information structure in the way we observed. In particular, why does contextual probability only impact the realization of corrective focus, while lexical frequency only impacts the realization of new-information focus? We suggest that these differential effects can be explained in terms of *how* ‘surprising’ the preceding utterance (the question asked by the partner) is to the participant—what we call **epistemic surprisal**. In a conversation, people might have expectation about what others might say, and when other people’s utterances contradict their expectation, they mark this surprisal prosodically in their responses.

Let us consider the low-probability corrective focus condition. Here, the correct information (the sentence read by the participant to correct the partner’s utterance) is improbable (e.g., *Bonnie and Laura found balls in the sea*). Relative to this correct ‘state of the world’, the partner’s misbelief (*Bonnie and Laura found boats in the sea*) may not strike the participant as surprising—because boats are quite likely to occur in the sea. This may be what motivates low pitch prominence (so low that it is comparable to the pitch in wide focus) in the participant’s response. On the other hand, when the correct information/‘state of the world’ is highly probable (e.g., *Bonnie and Laura found fish in the sea*), it may be quite surprising to the participant that the partner is mistaken about it—thus resulting in increased pitch prominence in the correction produced by the participant.

Possibly for similar reasons, new information conveyed by infrequent words is realized with low pitch prominence. In the new-information focus condition, if the new information is a relatively infrequent word (e.g., *shells*), it may not be surprising to the participant that the partner asks for this information (what was found in the sea). In contrast, when the new information is a higher-frequency, often-used word (e.g., *fish*), it may seem to the participant that the partner’s question is not worth asking, and thus surprising—triggering increased pitch prominence in the participant’s response. In general terms, the idea is that participants produce increased prosodic prominence in their response if the other person’s preceding utterance is surprising.

If these ideas are on the right track, our findings can be connected to the notion of ‘surprise’ or ‘astonishment’ developed by researchers in intonation and social interaction (e.g., Dombrowski, 2003, Local 1996, Niebuhr and Zellers 2012, Selting 1996, Wilkinson and Kitzinger 2006). Crucially, existing work distinguishes “the social expression of surprise (the public display of finding something counter to expectation)” from “the psychology of surprise (the emotional experience of encountering the unexpected)” (Wilkinson and Kitzinger 2006:152). Related to the social expression of surprise, so-called linguistic astonishment “is oriented towards the propositional content of the speaker’s message. It relates a local piece of information to the overall argumentation by denoting a factual mismatch between the speaker and the external world.” (Niebuhr and Zellers 2012:160). Our thinking builds on these ideas:

Specifically, we use **epistemic surprisal** to refer to *the degree of difference between the speaker’s expectation and reality*. ‘Reality’ as used here is not limited to events or concrete objects, since a speaker can also be surprised by the preceding content of the conversation (e.g., what their partner just said). A higher degree of epistemic surprisal means a bigger mismatch between what has just happened/what has been said and what the speaker had expected (cf. Bayes’ rule). We use the term ‘epistemic surprisal’ to avoid confusing linguistic astonishment with emotional astonishment, and to indicate a conceptual connection to the information-theoretic definition of surprisal (e.g., Levy 2011, Hale 2003).

Our results can thus be recast as follows: *The higher the degree of epistemic surprisal experienced by the speaker, the higher the degree of prosodic prominence on the relevant words. We*

would like to suggest that people's expectations about their conversational partner's utterances are an important factor involved in the prosodic encoding of informativity; this may be the motivation behind the interaction between information-theoretic factors and information structure. Future work will allow us to further test the validity of this account.

5 Conclusions

We conducted a production study that investigated how two types of informativity factors interact in shaping the prosody of utterances: (i) information-theoretic properties, specifically word frequency and contextual probability, and (ii) information-structural categories, specifically new-information focus and corrective focus. Our results show that word frequency modulates the prosodic prominence associated with new-information focus, whereas contextual probability modulates the prosodic prominence resulting from corrective focus. With regard to pitch cues, new-information narrow focus is more prominent than wide focus only when the word carrying new information is lexically frequent, while corrective narrow focus is more prominent than wide focus only when the word conveying corrective information is contextually probable. We discuss how these patterns can be explained in terms of 'epistemic surprisal'—the extent to which the speaker's expectation differs from the reality. The observed interaction between information-theoretic factors and information structure might be motivated by this third factor: Speakers produce greater prosodic prominence when they are surprised at their conversational partner's utterances. By integrating informativity factors that have previously been examined in separate traditions of work, this study brings new insights into the prosodic encoding of informativity.

References

- Aylett, Matthew, and 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* 47:31–56.
- Baker, Rachel E., and Ann R. Bradlow. 2009. Variability in word duration as a function of probability, speech style, and prosody. *Language and Speech* 52:391–413.
- Bell, Alan, Jason M. Brenier, Michelle Gregory, Cynthia Girand, and Dan Jurafsky. 2009. Predictability effects on durations of content and function words in conversational English. *Journal of Memory and Language* 60:92–111.
- Bell, Alan, Dan Jurafsky, Eric Fosler-Lussier, Cynthia Girand, Michelle Gregory, and Daniel Gildea. 2003. Effects of disfluencies, predictability, and utterance position on word form variation in English conversation. *Journal of the Acoustical Society of America* 113:1001–1024.
- Breen, Mara, Evelina Fedorenko, Michael Wagner, and Edward Gibson. 2010. Acoustic correlates of information structure. *Language and Cognitive Processes* 25:1044–1098.
- Brysbart, Marc, and Boris New. 2009. Moving beyond Kucera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods* 41:977–990.
- Chen, Yuhui, and Bettina Braun. 2006. Prosodic realization in information structure categories in standard Chinese. In *Speech Prosody 2006*, ed. R. Hoffmann and H. Mixdorff. Dresden: TUD Press.
- Cooper, William E., Stephen J. Eady, and Pamela R. Mueller. 1985. Acoustical aspects of contrastive stress in question-answer contexts. *Journal of Acoustical Society of America* 77:2142–2156.
- Couper-Kuhlen, Elizabeth. 1984. A new look at contrastive intonation. In *Modes of Interpretation: Essays Presented to Ernst Leisi on the Occasion of His 65th Birthday*, ed. R. Watts and U. Weidman. Tübingen: Gunter Narr Verlag.
- Dombrowski, Ernst. 2003. Semantic features of accent contours: Effects of F0 peak position and F0 time shape. In *Proceedings of the 15th International Congress of Phonetic Sciences (ICPhS 2003)*, 1217–1220.
- Dik, Simon C. 1997. *The Theory of Functional Grammar*. Berlin: Mouton De Gruyter.
- Eady, Stephen J., William E. Cooper, Gayle V. Klouda, Pamela R. Mueller, and Dan W. Lotts. 1986. Acoustical characteristics of sentential focus: Narrow vs. broad and single vs. dual focus environments. *Language and Speech* 29:233–251.
- Garcia, Damien. 2010. Robust smoothing of gridded data in one and higher dimensions with missing values. *Computational Statistics and Data Analysis* 54:1167–1178.
- Gu, Chong. 2003. *Smoothing Spline ANOVA Models*. New York: Springer.

- Gu, Chong. 2013. Package gss: General Smoothing Splines. URL <http://cran.r-project.org/web/packages/gss/>.
- Hale, John. 2003. The information conveyed by words in sentences. *Journal of Psycholinguistic Research* 32:101–123.
- Hwang, Hyun Kyung. 2012. Asymmetries between production, perception and comprehension of focus types in Japanese. In *Proceedings of Speech Prosody 2012*, 326–329.
- Katz, Jonah, and Elisabeth Selkirk. 2011. Contrastive focus vs. discourse-new: Evidence from phonetic prominence in English. *Language* 87:771–816.
- Kawahara, Hideki, Alain de Cheveigne, and Roy D. Patterson. 1998. An instantaneous-frequency-based pitch extraction method for high-quality speech transformation: Revised TEMPO in the STRAIGHT-suite. In *Proceedings of the 5th International Conference on Spoken Language Processing (ICSLP '98)*, 1367–1370.
- Krahmer, Emiel, and Marc Swerts. 2001. On the alleged existence of contrastive accents, *Speech Communication* 34:391–405.
- Levy, Roger. 2011. Probabilistic linguistic expectations, uncertain input, and implications for eye movements in reading. *Studies of Psychology and Behavior* 9:52–63.
- Local, John. 1996. Conversational phonetics: Some aspects of news receipts in everyday talk. *Studies in Inter-actational Sociolinguistics* 12:177–230.
- Munson, Benjamin, and Nancy Pearl Solomon. 2004. The influence of phonological neighborhood density on vowel articulation. *Journal of Speech, Language, and Hearing Research* 47:1048–1058.
- Niebuhr, Oliver, and Margaret K. Zellers. 2013. Late pitch accents in hat and dip intonation patterns. In *Understanding Prosody—The Role of Context, Function and Communication*, ed. O. Niebuhr, 159–186. Berlin: Walter de Gruyter.
- Nolan, Francis. 2003. Intonational equivalence: an experimental evaluation of pitch scales. In *Proceedings of the 15th International Congress of Phonetic Sciences*, 771–774.
- Ouyang, Iris Chuoying, and Elsi Kaiser. 2013. Prosody and information structure in a tone language: An investigation of Mandarin Chinese. *Language and Cognitive Processes*. DOI 10.1080/01690965.2013.805795.
- Ouyang, Iris Chuoying, and Elsi Kaiser. 2014. Prosodic encoding of informativity: Word frequency and contextual probability interact with information structure. In *Proceedings of the 36th Annual Meeting of the Cognitive Science Society 2014 (CogSci 2014)*, 1120–1125.
- Pan, Shimei, and Julia Hirschberg. 2000. Modeling local context for pitch accent prediction. In *Proceedings of the 38th Annual Conference of the Association for Computational Linguistics (ACL-00)*, 233–240.
- Pluymaekers, Mark, Mirjam Ernestus, and R. Harald Baayen. 2005. Lexical frequency and acoustic reduction in spoken Dutch. *Journal of the Acoustical Society of America* 118:2561–2569.
- Schwarzschild, Roger. 1999. Givenness, AVOIDF and other constraints on the placement of accent. *Natural Language Semantics* 7:141–177.
- Selting, Margret. 1996. Prosody as an activity-type distinctive cue in conversation: The case of so-called ‘astonished’. In *Prosody in Conversation: Interactional Studies*, ed. E. Couper-Kuhlen and M. Selting, 231–270.
- Shue, Yen-Liang, Patricia Keating, Chad Vicenik, and Kristine Yu. 2011. VoiceSauce: A program for voice analysis. In *Proceedings of the 17th International Congress of Phonetic Sciences (ICPhS XVII)*, 1846–1849.
- Vallduví, Enric, and Maria Vilkuna. 1998. On rheme and contrast. In *Syntax and Semantics 29: The Limits of Syntax*, ed. P.W. Culicover and L. McNally. San Diego: Academic Press.
- van Son, R.J.J.H., Florian J. Koopmans-van Beinum, and Louis C.W. Pols. 1998. Efficiency as an organizing principle of natural speech. In *Proceedings of the 5th International Conference on Spoken Language Processing (ICSLP '98)*, 2375–2378.
- Wagner, Michael, and Duane G. Watson. 2010. Experimental and theoretical advances in prosody: A review. *Language and Cognitive Processes* 25:905–945.
- Wilkinson, Sue, and Celia Kitzinger. 2006. Surprise as an interactional achievement: Reaction tokens in conversation. *Social psychology quarterly* 69:150–182.
- Wright, Richard. 2003. Factors of lexical competition in vowel articulation. In *Phonetic Interpretation: Papers in Laboratory Phonology VI*, ed. J. Local, R. Ogden and R. Temple. Cambridge: Cambridge University Press.

Department of Linguistics
 University of Southern California
 Los Angeles, CA 90089
 chuoyino@usc.edu
 emkaiser@usc.edu