Language and Speech

Article

Language *and* Speech
2017, Vol. 60(1) 123–153

DOI: 10.1177/0023830916647079
journals.sagepub.com/home/las
SAGE

The Role of Predictability in Intonational Variability

Rory Turnbull
Ohio State University, USA

Abstract

Predictability is known to affect many properties of speech production. In particular, it has been observed that highly predictable elements (words, syllables) are produced with less phonetic prominence (shorter duration, less peripheral vowels) than less predictable elements. This tendency has been proposed to be a general property of language. This paper examines whether predictability is correlated with fundamental frequency (F0) production, through analysis of experimental corpora of American English. Predictability was variously defined as discourse mention, utterance probability, and semantic focus. The results revealed consistent effects of utterance probability and semantic focus on F0, in the expected direction: less predictable words were produced with a higher F0 than more predictable words. However, no effect of discourse mention was observed. These results provide further empirical support for the generalization that phonetic prominence is inversely related to linguistic predictability. In addition, the divergent results for different predictability measures suggests that the parameterization of predictability within a particular experimental design can have significant impact on the interpretation of results, and that it cannot be assumed that two measures necessarily reflect the same cognitive reality.

Keywords

Predictability, fundamental frequency, intonation

1 Introduction

According to the Smooth Signal Redundancy Hypothesis (SSRH) (Aylett, 2000; Aylett & Turk, 2004, 2006; Turk, 2010), prosodic prominence is inversely correlated with contextual predictability. This hypothesis has been supported by a diverse array of evidence suggesting that phonetic reduction is inversely correlated with predictability: more frequent words are produced with shorter duration (Gahl, 2008), more contextually probable syllables have a shorter duration (Aylett & Turk, 2004), and words are less intelligible upon their second mention in a discourse (Fowler & Housum, 1987). However, most of these studies have been limited to aspects of segmental realization and

Corresponding author:
Rory Turnbull, Laboratoire de Sciences Cognitives et Psycholinguistique, École Normale Supérieure, 29 Rue d'Ulm, 75005, Paris, France
Email: rory.turnbull@ens.fr

timing, such as duration or vowel peripherality, and have not examined the prosodically relevant dimension of fundamental frequency (F0). This paper fills a gap in the empirical coverage of predictability effects through an examination of the role of different dimensions of predictability in realization of F0, using duration as a point of comparison.

This investigation is important since it addresses one of the direct predictions of the SSRH. While some discussions of the SSRH (and related theories) have characterized the theory in terms of its predictions about duration and segmental timing, it is worth noting that the SSRH's conception of "prosodic prominence" is not purely temporal. Indeed, both Aylett (2000) and Turk (2010) are quite explicit in their definitions of prosodic prominence as being ultimately derived from acoustic salience. This notion of the relevance of non-temporal cues is not purely abstract: Aylett (2000) and Aylett & Turk (2006) examined the non-temporal cue of vowel peripherality and found that more predictable vowels tend to be less peripheral in the vowel space, relative to less predictable vowels. The SSRH has therefore always had a role for non-temporal cues, such as F0.

The data examined here are drawn from two experimental corpora. These corpora are from experiments involving a naïve participant directing a confederate in a goal-oriented and cooperative communicative task. Predictability can be operationalized within an experimental context in a number of different ways. In the present analysis, three different conceptions of contextual predictability are examined—discourse mention, utterance probability, and semantic focus. These variables are examined with a view toward evaluating the hypothesis that more extreme F0 values, that is, higher peaks, are observed on less predictable words than on more predictable words. Since F0 varies as a consequence of phonological prosodic structure, these structures must be taken into account before an analysis of predictability can be attempted. These sources of F0 variability are discussed in the following section.

1.1 F0 variation

Phonological variation in F0 has been fruitfully described under the autosegmental-metrical framework of intonational phonology (Beckman & Pierrehumbert, 1986; Ladd, 2008; Pierrehumbert, 1980). Under this framework, pitch movements are a reflection of sequences of underlying tones at the postlexical level, just as segmental acoustic variation (such as shifts in F2) are thought to derive from sequences of phonemes. The majority of contemporary work on American English intonation within the autosegmental-metrical framework uses the Tones and Break Indices (ToBI) annotation system (Beckman & Hirschberg, 1994; Silverman et al, 1992). This system, for American English, involves three classes of tones: pitch accents, which associate with stressed syllables; phrasal tones, which associate with the right edge of smaller prosodic phrases; and boundary tones, which associate with the right edge of larger prosodic phrases. ToBI also marks break indices, which indicate the relative strength of prosodic breaks between words.

It has been claimed that different types of pitch accents are used to mark different discourse functions (Pierrehumbert & Hirschberg, 1990; Pierrehumbert & Steele, 1990), although the relationship is not necessarily one-to-one (Ito, Speer, & Beckman, 2004). Assuming that different pitch accents are realized with different F0 (Beckman & Pierrehumbert, 1986), and that different pitch accents relate to information structures, it follows that there is an expected relationship between information structures and F0, a prediction borne out by decades of instrumental research (e.g., Cooper, Eady, & Mueller, 1985; Eady & Cooper, 1986; Eady, Cooper, Klouda, Mueller, & Lotts, 1986; Face, 2001; Xu, 1999; Xu & Xu, 2005). However, the SSRH predicts a much more finely grained relationship between F0 and predictability. More specifically, the SSRH predicts that F0 peaks should be less prominent (i.e., lower) on more predictable elements. Put another way, the purpose of this study is to determine, after controlling for phonological effects of pitch accent, whether peak F0 is negatively correlated with predictability.

One study investigating the role of expectedness in prosodic realization (Watson, Arnold, & Tanenhaus, 2008) found tentative support for this hypothesis—words in the unpredictable context were produced with overall higher F0 than words in the predictable context. However, the predictability manipulation was binary rather than continuous, and the analysis did not control for potential effects of pitch accenting. The current study builds on this work through analysis of speech corpora that allow for continuous and non-binary definitions of predictability and control of pitch accent.

1.2 Utterance probability

In one of the earliest studies on predictability effects in speech, Lieberman (1963) demonstrated that words which were less probable given the preceding context were more intelligible in isolation than the same words in a more probable context. The classic example is the word *nine* in the sentences *a stitch in time saves nine* and *the number that you will hear is nine*, where the target word is more predictable in the former but less predictable in the latter. Subsequent research, drawing heavily on advances in statistical natural language processing, has refined the definitions and quantifications of word probability (Bell, Brenier, Gregory, Girand, & Jurafsky, 2003; Bell, Jurafsky, Fosler-Lussier, Girand, Gregory, 2009; Jurafsky, Bell, Gregory, & Raymond, 2001; Pierrehumbert, 2003). These studies, among others, have helped firmly establish Lieberman's (1963) basic generalization: less probable words are produced with greater acoustic prominence than more probable words.

Predictability is conceptually related to accessibility, which is the extent to which a referent is attended to within a discourse (Chafe, 1974; Grosz & Sidner, 1986). However, it should be noted that these concepts are formally distinct and can be orthogonal, since the latter relates to the behavior of the interlocutors, while the former is a reflection of statistical facts about the world (see also Wagner & Klassen, 2015). While the relationship between accessibility and prosodic prominence has received much recent attention (e.g., Baumann & Riester, 2012; Dahan, Tanenhaus, & Chambers, 2002; *inter alia*), this paper does not directly address accessibility.

Although the terms 'predictability' and 'probability' are often used interchangeably, in this paper I wish to draw a sharp and meaningful distinction between them. By 'probability' I refer to the mathematical formulation of the likelihood of the occurrence of an event, calculated as a ratio. Probability is thereby directly computable in a given context, and does not necessarily accord with human intuition (e.g., the (in)famous Monty Hall problem or the base rate fallacy (Lyon & Slovic, 1976)). On the other hand, 'predictability' refers to a cognitive status, the (potentially subconsciously) perceived chance of an event, which is inaccessible to direct examination (see also Kahneman & Tversky, 1982). This relation between predictability and probability is conceptually similar to the distinction between pitch and F0, competence and performance, or loudness and amplitude, respectively, and is conceptually similar to Kahneman & Tversky's (1972) "subjective probability". Utterance probability is one of several ways by which predictability can be approximated or examined.

1.3 Discourse mention

Words tend to be phonetically reduced on their second or subsequent mention; this phenomenon is referred to in the literature as second mention reduction (Baker & Bradlow, 2009; Fowler & Housum, 1987). Second mention reduction has traditionally been treated as a semantically motivated effect related to givenness, bolstered by findings such as those of Bard, Lowe, and Altmann (1989), who found that second mentions failed to reduce if the word had a different referent than the first mention. However, more recent research (such as Baker and Bradlow, 2009) has uncovered second mention reduction on nouns with dissimilar referents, adjectives, and verbs, suggesting that a purely discourse-semantic-driven account is inadequate. These findings are consistent with models where mentioning a word in discourse primes access at the lexical rather than the semantic level (e.g., Bard et al., 2000; Bell et al., 2009; Lam & Watson, 2014).

This phenomenon has been attributed to the fact that first mentions tend to be discourse-new and thus referentially unpredictable, while second (and subsequent) mentions tend to be discourse-given and referentially predictable (Kahn & Arnold, 2012). Following this line of reasoning, discourse mention is thus a type of predictability (Pluymaekers, Ernestus, & Baayen, 2005).

A connection is purported to exist between givenness and pitch accenting, such that discourse-given words tend to be unaccented and discourse-new words and words in focus should be accented (Chafe, 1974; Féry & Samek-Lodovici, 2006; Krahmer & Swerts, 2001; Terken & Hirschberg, 1994), but Calhoun (2010a) presented evidence for a more explicitly nuanced view where the relationship between accenting and givenness is probabilistic and bound by metrical constraints. Nonetheless, after controlling for pitch accenting, the SSRH predicts that subsequent mentions of words will be produced with a lower F0 than first mentions.

1.4 Focus

Many studies of F0 have looked at effects of focus, also termed “contrastive focus”, “rheme”, or “comment” with often slightly different definitions.[1] Such research has generally been carried out under the umbrella of investigation into the relationship between information structure and prosody. However, I argue that focus, a semantic property, is inherently a type of predictability. According to Schwarzschild (1999), all non-focused material is assumed to be contextually given. Although the states of ‘given’ and ‘new’ are often presented as a binary dichotomy, a more complex schema may better reflect linguistic processing, either a discrete hierarchy as proposed by Gundel, Hedberg, and Zacharski (1993), or a continuum of predictability or accessibility (Birner, 1994). Therefore, all else being equal, material in focus is less predictable than material not in focus, since material not in focus must be given, while material in focus can vary in information structure. (It is important to bear in mind that focus is properly considered a semantic property and not a prosodic property.) This conclusion is further supported by work in natural language processing which successfully uses predictability information to automatically determine givenness (e.g., Rahman & Ng, 2011). It is clear, then, that predictability is one component of focus. We expect that words in focus will have a higher F0 than words not in focus.

1.5 Predictions

To summarize the foregoing discussion, the SSRH predicts that prosodic prominence should be inversely proportional to linguistic predictability. Utterance predictability, discourse mention, and semantic focus are three linguistic properties of which predictability is an aspect. We expect that words with a higher utterance probability, words that are subsequent mentions, and words that are in focus will be realized with a higher F0 than words with a lower utterance probability, words that are first mentions, or words that are not in focus.

However, we must also acknowledge that things may not be so simple. While words are linearly organized in speech, they possess hierarchical structure. This structure means that words can vary in their syntactic function (e.g., a word either is or is not the head of a phrase) and their lexical specification (e.g., a word can be an adjective or a noun), independently of linear position. To assume that hierarchical structure has no influence upon predictability would be naïve.

In addition, the interplay among factors contributing to linguistic predictability is not necessarily straightforward. While it is neat for linguistic theories if effects are orthogonal and additive, real data often exhibit complex interactions, subadditive effects, superadditive effects, and nonlinear structures. Part of this complexity may arise as a consequence of the analysis frames that linguists impose upon the phenomena investigated, but some of it could equally be due to the underlying processes being genuinely complex. In any event, the basic predictions made by the SSRH are quite straightforward, but the details around how these factors play out are quite unspecified.

2 Experiment 1—Christmas tree decoration

2.1 Introduction

The first corpus comes from an experiment, described in Ito and Speer (2006), which involved a communicative task designed to elicit spontaneous, natural dialog. The corpus is ToBI-annotated, and the structure of the experiment allows for examination of effects of discourse mention and focus.

2.2 Corpus

In this experiment, one person (a naïve participant) instructed another (a confederate) to decorate a miniature Christmas tree with ornaments. The participant could see the tree being decorated, but not the confederate or the available ornaments. The ornaments were distinguished by both color and shape (e.g., there was a blue bell, a green angel, a green bell, etc). Table 1 lists all of the colors and shapes used in the experiment.

In each trial, a picture appeared on a computer monitor in front of the participant. The picture illustrated which ornament was to be selected, and where on the tree it was to be placed. A description of the ornament was also provided (e.g., *blue drum*); see Figure 1 for an example display. The participant conveyed this information to the confederate, who hung the ornament as directed. The participant was given no instructions on how to speak—that is, their productions were spontaneous and entirely free from task-specific constraints. The confederate was trained to never utter the target words (e.g., *blue drum*), but to instead use anaphora (e.g., *this one*).

The order of the ornaments presented on screen was manipulated to create contexts whereby one linguistic element is in focus. For instance, in a sequence of trials involving first a *blue drum* followed by a *green drum*, the second utterance bears focus on the word *green*, and this focus is expected to be manifested in the signal (through, for example, a L+H* pitch accent). The order was manipulated to create three possible focus conditions: focus on the adjective, focus on the noun, and focus on the entire noun phrase (NP).[2]

Table 1. Adjectives and nouns used in the Christmas tree decoration task.

Adjective	Noun
blue	ball
green	egg
orange	onion
grey	house
beige	bell
navy	drum
brown	doll
clear	candy

Figure 1. Example monitor display for the naïve participant from Ito and Speer's (2006) Christmas tree decoration task.

As the experiment continued, there was repetition of particular adjectives or of particular nouns, but each unique adjective–noun combination was a target ornament only once. For example, in the course of decorating a tree, the word *blue* was used three times to refer to distinct blue ornaments; but the phrase *blue drum* was a target utterance only once. Each adjective or noun was used a maximum of three times in the decoration of a particular tree. There were thus two manipulations for the words in a given trial: the focus condition (NP focus, adjective focus, or noun focus) and the number of previous mentions of the words in the target phrase.

2.3 Acoustic measurements

Here, data from nine native speaker participants were analyzed, using the ToBI annotations provided in Ito and Speer's (2006) dataset. Ito and Speer's ToBI annotations were carried out by at least two trained annotators (see also Ito et al., 2004). Measurements of the stressed vowel's F0 maximum height in Hz were taken for each target word, using a frame duration of 10ms and Boersma's (1993) autocorrelation algorithm. Given the role that predictability plays in temporal variability, vowel duration in milliseconds was also measured, to be used as a point of comparison for the F0 measurements. F0 maxima and vowel boundaries were identified through visual and auditory inspection of the waveform, spectrogram, and automatic F0 trace. Peaks outside of the stressed vowel were marked if such a delayed peak was consistent with the phonological annotation (e.g., a L*+H pitch accent). A word was defined as a target word if it was in a two-word phrase consisting of an adjective and a noun from Table 1, uttered in reference to the current item. This definition excludes cases where the participants used alternative lexical items, such as "disco ball", or more complex phrases, such as "the clear doll cupid thing". Also excluded by this definition are phrases which referred to previously placed ornaments, such as "underneath the orange drum", since these utterances do not refer to the target item.

2.4 Data analysis

The ToBI-annotated pitch accents were collapsed into three categories for ease of analysis: rising (L+H* and L*+H), high (H* and !H*), and unaccented (words with no pitch accent), following other analyses of similar corpora (Burdin & Clopper, 2015; Burdin, Phillips-Bourass, Turnbull, Yasavul, Clopper, & Tonhauser, 2015; Turnbull, Burdin, Clopper, and Tonhauser, 2015). The modeling procedure involved fixed effects of adjective pitch accent (rising, high (baseline), or unaccented), noun pitch accent (rising, high (baseline), or unaccented), focus condition (adjective, noun, or NP (baseline)), number of adjective mentions (from 1 to 3), number of noun mentions (from 1 to 3), and all possible two-way interactions.[3] Random intercepts of participant, adjective, and noun were included.[4] All continuous variables were centered at zero prior to being entered into the model.

The data were analyzed using stepwise mixed-effect regression modeling to predict the F0 maxima and vowel duration of the target adjective and the target noun. Adjective and noun were modeled separately since, although the overall predictions for each word are the same (more acoustic prominence on less predictable words), adjectives and nouns differ in a number of linguistically relevant ways which could influence the magnitude of observed effects. Principal among these difference is syntax; the phrase-final noun is the head of the syntactic phrase (the NP), while the adjective is an adjunct.

Model selection for the stepwise mixed-effects regression was performed using a best-path backwards algorithm. Beginning with the maximal model structure, each fixed effect was evaluated for exclusion from the model via a likelihood-ratio test. If an effect did not contribute to overall data likelihood given the model (i.e., $p_{\chi^2} \geqslant 0.05$), it was removed. If more than one effect met this criterion, the effect with the highest *p*-value was removed. This process was then repeated iteratively with the new, smaller model, until no more exclusions were possible—that is, until all effects contributed to data likelihood at $p_{\chi^2} < 0.05$. The algorithm operated under the constraint that an interaction term could only exist in a model structure if all components of the interaction were also present as main effects. This constraint makes the final model easier to interpret and reduces the chance of overfitting the model. An additional constraint operated on the algorithm. For models where a categorical factor was involved in an interaction term, such as A + B + (A × B) where A is a categorical factor, the log-likelihood tests were carried out on models without the interaction. That is, to assess the significance of factor A in the above model, the model A + B would be compared with A. This method is due to the computational representation of categorical factors making A + (A × B) functionally equivalent to A + B + (A × B). Finally, for categorical factors with more than two levels, log-likelihood testing will reveal only if there is an overall effect of the factor, but not of which levels are significantly different from which other levels. (This result is conceptually equivalent to a 'main effect' in an ANOVA model.) To assess significance between the factor levels, the *t*-statistic associated with each coefficient was examined, and absolute values greater than 2 were considered significant (Baayen, Davidson, & Bates, 2008; Barr, Levy, Scheepers, & Tily, 2013).

2.5 Results

A total of 849 target NPs consisting of an adjective followed by a noun were extracted from the corpus. Of those phrases, F0 could not be reliably extracted from 25 adjectives and 62 nouns, due to non-modal voicing quality, noise in the recording, or other concerns. In addition, any measurement more than three standard deviations from the subject mean was excluded. This criterion led to the exclusion of eight adjective F0 measures, seven noun F0 measures, eight adjective vowel duration measures, and seven noun vowel duration measures. Therefore, the final data set analyzed consisted of 816 adjective F0 measurements, 780 noun F0 measurements, 841 adjective vowel duration measurements, and 842 noun vowel duration measurements.

Table 2 summarizes the pitch accents observed on the adjectives and nouns, split by focus condition. Note that the pitch accents vary considerably with focus condition—for example, in the noun

Table 2. Counts (and percentages) of pitch accent types on adjectives and nouns in the Christmas tree decoration experiment, split by focus condition.

Focus condition	High		Unaccented		Rising		Total	
Adjectives:								
NP focus	231	(80.49)	3	(1.05)	53	(18.47)	287	(100)
Adjective focus	111	(38.81)	1	(0.35)	174	(60.84)	286	(100)
Noun focus	209	(76.28)	15	(5.47)	50	(18.25)	274	(100)
Nouns:								
NP focus	221	(77.00)	21	(7.32)	45	(15.68)	287	(100)
Adjective focus	130	(45.45)	135	(47.20)	21	(7.34)	286	(100)
Noun focus	172	(62.77)	14	(5.11)	88	(32.12)	274	(100)

Table 3. Tables of coefficients and log-likelihood comparisons for each retained fixed effect, adjective F0 peak model in the Christmas tree decoration experiment.

Variable	β	*SE*	*t*
Intercept	−1.384	17.696	−0.078
Adjective: no accent	−11.627	5.545	−2.097
Adjective: rising accent	6.066	1.962	3.092
Noun: no accent	3.359	2.260	1.486
Noun: rising accent	−5.828	2.111	−2.761
Variable	χ^2	*DF*	*p*
Adjective pitch accent	23.904	2	<.001
Noun pitch accent	13.878	2	<.001

focus condition, nouns are rarely unaccented (7%), but they are unaccented in a plurality of cases (47%) in the adjective focus condition. This uneven distribution of pitch accents is relevant is relevant insofar as it suggests an intimate and complex relationship between focus, pitch accent, and F0.

2.5.1 Adjective F0 maximum. As Table 3 shows, two main effects remained in the model of adjective F0 maximum: the adjective pitch accent type, and the noun pitch accent type. The model revealed that adjectives with a rising pitch accent were realized with a higher peak F0 ($M = 175$ Hz, $SD = 56.1$ Hz) than adjectives with a high pitch accent ($M = 168$ Hz, $SD = 56.7$ Hz), which were in turn significantly higher than unaccented adjectives ($M = 158$ Hz, $SD = 58.9$ Hz). The means and standard deviations reported here are calculated over subject means (i.e., $N = 9$), to abstract away from differences between participants in their proportions of pitch accents used. All means and standard deviations in this paper are calculated in this way; the standard deviations are best interpreted as a measure of variance between subjects. The effect whereby rising accents have higher peak F0 than high accents, which in turn have higher peak F0 than unaccented words, is consistent with previous research on the F0 of words with different pitch accents (e.g., Calhoun, 2012).

The peak F0 on the adjective was also found to vary with the pitch accent type of the noun, that is, the following word in the phrase. When the following noun bore a rising accent, a significantly lower F0 was observed on the adjective ($M = 165$ Hz, $SD = 54.0$ Hz), compared to when the following noun was high ($M = 171$ Hz, $SD = 56.0$ Hz) or unaccented ($M = 175$ Hz, $SD = 54.9$ Hz). This effect is consistent with the notion that prosodic prominence, implemented via F0, is

Table 4. Tables of coefficients and log-likelihood comparisons for each retained fixed effect, noun F0 peak model in the Christmas tree decoration experiment.

Variable	β	*SE*	*t*
Intercept	−3.399	14.720	−0.231
Noun: no accent	−9.829	3.176	−3.095
Noun: rising accent	6.750	2.775	2.433
Adjective: no accent	3.172	6.948	0.456
Adjective: rising accent	−3.556	2.718	−1.308
Focus: Adjective	9.648	2.724	3.542
Focus: Noun	3.301	2.390	1.381
Variable	χ^2	*DF*	*p*
Noun pitch accent	20.396	2	<.001
Adjective pitch accent	20.249	2	<.001
Focus condition	12.444	2	.002

inherently relative: a word is only prominent insofar as it is distinct from other words (Kohler, 2008; Turnbull, Royer, Ito, & Speer, 2014). In addition, this effect speaks to the notion that phonetic implementation of pitch accents is not mere interpolation between acoustic targets. Rather, coarticulatory forces affect F0 realization as they do other phonetic correlates of phonological structure, such as F2 (Grabe, 1998; Ladd & Schepman, 2003).

2.5.2 Noun F0 maximum. In the noun F0 maximum model, three fixed effects were retained, as seen in Table 4: noun pitch accent type, adjective pitch accent type, and focus condition. As expected, nouns with a rising pitch accent had a higher F0 (M= 161 Hz, SD = 45.8 Hz) than nouns with a high pitch accent (M = 155 Hz, SD = 47.0 Hz), which in turn were higher than unaccented nouns (M = 150 Hz, SD = 51.3 Hz). The fixed effect of adjective pitch accent (i.e., the preceding word) significantly contributed to the data likelihood, although the individual effects were not significant (both $t < 2$). This effect mirrors the effect of noun pitch accent on adjective F0, and underscores the observation that the F0 of pitch accents is crucially dependent upon the broader context in which they occur.

In addition, an effect of focus was observed such that nouns in adjective focus phrases had a significantly higher peak F0 (M = 155 Hz, SD = 50.3 Hz) than nouns in phrases with NP focus (M = 152 Hz, SD = 44.8 Hz). Interestingly, nouns in noun focus phrases had even higher F0 peaks (M = 157 Hz, SD = 47.2 Hz), but this effect was not significantly higher than the NP focus condition. This result is unexpected—prior research (e.g., Krahmer & Swerts, 2001) suggests that the noun should have approximately equally high peaks in both the NP focus condition and the adjective focus condition, but higher peaks in the noun focus condition.

2.5.3 Adjective vowel duration. Two fixed effects were retained in the adjective vowel duration model: adjective pitch accent type and number of adjective mentions. Table 5 provides the model summaries. The individual effects of pitch accent type were not significant, but the fixed effect contributed significantly to the model. The vowels of accented adjectives (M = 150 ms, SD = 15.8 ms) tended to be longer than the vowels of unaccented adjectives (M = 127 ms, SD = 32.5 ms). This finding is consistent with previous work suggesting that unaccented syllables tend to be shorter than accented syllables (Cambier-Langeveld & Turk, 1999; Turk & Sawusch, 1997).

Table 5. Tables of coefficients and log-likelihood comparisons for each retained fixed effect, adjective duration model in the Christmas tree decoration experiment.

Variable	*β*	*SE*	*t*
Intercept	1.187	12.786	0.093
Adjective: no accent	−14.392	8.614	−1.671
Adjective: rising accent	−0.804	2.835	−0.284
Number of adjective mentions	−3.355	1.511	−2.220
Variable	χ^2	*DF*	*p*
Adjective pitch accent	23.034	2	<.001
Number of adjective mentions	4.915	1	.027

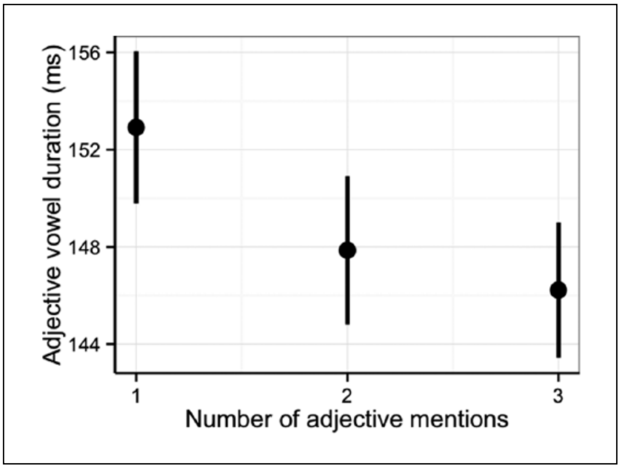

Figure 2. Means and standard errors of adjective vowel duration as a function of the number of adjective mentions.

As expected, second mention reduction was observed: the vowels of adjectives tended to be shorter on their second (*M* = 148 ms, *SD* = 17.2 ms) and third mentions (*M* = 146 ms, *SD* = 14.3 ms) relative to the first (*M* = 153 ms, *SD* = 17.4 ms). Figure 2 shows means and standard errors of the vowel duration at different numbers of mentions.

2.5.4 Noun vowel duration. Comparable results were obtained for the noun vowel duration model, which retained three fixed effects: noun pitch accent type, adjective pitch accent type, and number of adjective mentions. Table 6 provides the model summaries. As expected, unaccented nouns had significantly shorter vowels (*M* = 165 ms, *SD* = 46.7 ms) than vowels in nouns with high pitch accents (*M* = 181 ms, *SD* = 34.4 ms). Vowels in nouns with rising pitch accents were longer (*M* = 196 ms, *SD* = 50.0 ms) than vowels in nouns with high pitch accents, but not significantly so. The pitch accent of the adjective (the preceding word) contributed to data likelihood, but neither of the individual effects reached significance.

As expected, mention effects were observed, but the significant effect was that of *adjective* mentions. Noun vowels were shorter when the adjective had been mentioned more times. This effect suggests that the adjective mentions effect served to reduce the production of both the

Table 6. Tables of coefficients and log-likelihood comparisons for each retained fixed effect, noun duration model in the Christmas tree decoration experiment.

Variable	β	*SE*	*t*
Intercept	3.927	19.481	0.202
Noun: no accent	−18.557	4.671	−3.973
Noun: rising accent	8.020	4.387	1.828
Adjective: no accent	2.257	10.939	0.206
Adjective: rising accent	−4.589	4.058	−1.131
Number of adjective mentions	−4.915	1.925	−2.553
Variable	χ^2	*DF*	*p*
Noun pitch accent	26.131	2	<.001
Adjective pitch accent	20.476	2	<.001
Number of adjective mentions	6.494	1	.011

adjective *and* the noun. This 'spreading' of acoustic reduction is conceptually similar to Arnold, Kahn, and Pancani's (2012) finding of acoustic reduction on the determiner of a contextually predictable noun phrase. In that experiment, naïve participants instructed a confederate to move particular objects; the determiner of the referring noun phrase was observed to be produced with a shorter duration when it was clear that the confederate was anticipating the referent of the noun. That is, in the phrase *the teapot*, the determiner was reduced when the confederate was already holding the teapot in anticipation of the instruction. In the present data, we see a similar effect of the predictability of one element (the adjective) influencing the production of another (the noun). (Note, however that the Arnold et al. (2012) effect is of the predictability of a word affecting the production of the preceding word, while the present effect is that of the predictability of a word affecting the production of the following word.) No effect of noun mention on noun vowel duration was observed. Although the raw data suggest a trend (first mention mean vowel duration: 187 ms; subsequent mention mean vowel duration: 175 ms), the inclusion of the effect in the model did not significantly contribute to data likelihood.

2.6 Discussion

The results revealed expected effects of pitch accenting. Words with high and rising pitch accents were realized with higher F0 and longer duration than words without pitch accents, and rising pitch accents (like L+H* and L*+H) were realized with higher F0 than high pitch accents, consistent with previous research (e.g., Calhoun, 2012). However, the pitch accent of the adjective *and* the noun was needed in both of the F0 models to effectively predict F0. This result suggests that pitch accents are not invariant acoustic targets, but are subject to phonetic coarticulation effects in the same manner as segments (e.g., Öhman, 1966), and underscores the importance of considering F0 in context.

Effects of focus condition on F0 were not observed for the adjectives, and inconsistently observed for the noun. It would appear that the pitch accents accounted for the majority of the F0 variance related to focus; the pitch accenting factors were always present in the optimal F0 models. While the use of pitch accents in focus marking was not a question of the current analysis, the trends of use accord with expectations based on previous studies of accenting (Bolinger, 1972; Ito et al., 2004; Ito & Speer, 2006; Turnbull et al., 2014): in the NP focus condition, 199 (69.3%) of the phrases were realized with a high–high sequence; in the adjective focus condition, 113 (39.5%) of the phrases were realized with a rising–unaccented sequence, and 87 (30.4%) with a high–high sequence; and in the noun focus condition, 152 (55.5%) phrases were realized with a high–high sequence, and 52 (19.0%) with a high–rising sequence. From the point of view of contemporary theories of information structure, one aspect of these results is striking: non-focused material is not expected to bear a pitch accent, but here, many non-focused nouns (and some adjectives) are accented. Despite being at odds with the theoretical literature, however, these results are actually concordant with the rates of deaccenting in American English observed in experimental studies (Burdin et al, 2015; Calhoun, 2004, 2010b; Ito & Speer, 2006; Katz & Selkirk, 2011; Turnbull et al., 2015; see also Féry & Kügler, 2008 and Riester & Piontek, 2015 on German).

Similarly, focus was not observed to have an effect on vowel duration, a conclusion which appears to be at odds with evidence suggesting that focused elements are longer than unfocused elements (e.g., Eady & Cooper, 1986). However, as with the F0 patterns, a plausible interpretation of this lack of a significant effect is that the variability in duration is largely due to variation in pitch accent, since accented words tend to be longer than unaccented words (e.g., Burdin & Clopper, 2015; see also Tables 5 and 6).

A curious pattern that emerged from the results is the asymmetry of effects on the noun and the adjective. The realization of the noun appears to depend in large part on factors relating to the adjective, such as number of mentions, pitch accent, and focus condition, while the realization of the adjective does not depend on those same factors for the noun. Conceptually similar positional asymmetries have been reported in the literature. For example, it is well established that the duration of content words is related to the conditional probability of the word given the immediately following word (Bell et al., 2003, 2009; Seyfarth, 2014; Tily et al., 2009; Tily & Kuperman, 2012). Likewise, some theories of language production have proposed that word lengthening signals difficult or unpredictable upcoming material (e.g., Fox Tree & Clark, 1997). However, these results all implicate upcoming material in having an influence on word production, rather than previous material (in this case, the adjective) having an influence on word (the noun) production. Alternatively, Xu (1994) demonstrated that the F0 realization of Mandarin tones is influenced to a large degree by the immediately preceding tone, and that the following tone had comparatively little influence. Linear sequence, then, is an important variable for future investigation. Another possible factor that may help lead to an explanation is that the noun is in a privileged position, both in terms of syntax—it is the head of the phrase—and in terms of phrasing—it is phrase-final. These factors could plausibly have led to the observed asymmetry between nouns and adjectives. For example, it is known that mention effects are sensitive to high-level discourse structure (Fowler, Levy, & Brown, 1997), and they could also be sensitive to syntactic structure (e.g., Heller & Goldrick, 2014). Teasing apart the effects of syntactic prominence from positional prominence requires cross-linguistic investigation of languages with different NP structures—see for example Burdin et al. (2015) on English, Guaraní, K'iche', and Moroccan Arabic, or Swerts, Krahmer, and Avesani (2002) on Dutch and Italian.

As expected, second mention temporal reduction was observed on the adjectives. Second and subsequent mentions of adjectives were shorter in duration than first mentions. Mention was not observed to influence F0, implying that predictability, framed in terms of discourse mention, does not affect F0 production (see also Cole, Mo, & Hasegawa-Johnson, 2010). Mention is a relatively coarse measure, often couched in binary terms (Galati & Brennan, 2010), and it may not provide sufficient granularity to allow the observation of effects of predictability on F0. Utterance probability may provide the necessary granularity, and is therefore the focus of the next experimental corpus analysis.

Table 7. Adjectives and nouns used in the tile-placing experiment.

Adjective	Noun
blue	deer
green	owl
orange	flower
yellow	train
brown	lion

3 Experiment 2—Tile-placing task

3.1 Introduction

The second dataset is drawn from an experiment described in Burdin et al. (2015) and Turnbull et al. (2015). Like Ito and Speer's (2006) Christmas tree decoration experiment, this experiment involved a communicative task between a naïve participant and a confederate, but with more constraints on possible utterances. The experiment involved a manipulation of focus, and two manipulations of utterance predictability: probability of utterance content, and probability of utterance information structure. This experiment is therefore well suited to our investigation of the effects of predictability on F0 realization.

3.2 Corpus

Like the experiment of the first corpus, this experiment involved a naïve participant instructing a confederate; critical target phrases were adjective–noun sequences describing colored objects. In this experiment, the instructions were to place paper tiles depicting objects into numbered boxes on a game board. The tiles were distinguished both by color and by shape; the adjectives and nouns used to refer to those colors and shapes are listed in Table 7.

The participant was given a board with the end-state depicted, that is, all of the object tiles in all of the appropriate boxes; an example is shown in Figure 3. The confederate sat across a table with a blank board with no object tiles in any boxes. Object tiles were then placed on the table between the participant and the confederate. Both interlocutors could see each other and the available tiles, but the participant could not see the confederate's game board (and vice-versa). The participant was instructed to phrase all directions in the form of "put the adjective noun in box number". All participants began instruction with box one, and proceeded in order through the other boxes, although there was no implicit or explicit instruction to do so. The event of filling a board is referred to here as a trial; each trial therefore consisted of four or six utterances, depending on the size of the board.

The experiment consisted of four blocks. The first three blocks involved trials with four boxes and five available tiles. Each block consistently had the same constituent in focus—that is, one block was purely adjective focus trials, one block was purely noun focus trials, and another block was purely NP focus trials. In adjective focus trials, all of the available tiles had the same shape, and differed only in color. In noun focus trials, all of the available tiles had the same color, and differed only in shape. In NP focus trials, all of the available tiles were unique in both color and shape. See Burdin et al. (2015) for more detail on this manipulation and theoretical reasoning on why focus necessarily follows from this manipulation. These three blocks were dubbed 'predictable' trials, since which constituent was focused was predictable from the global context.

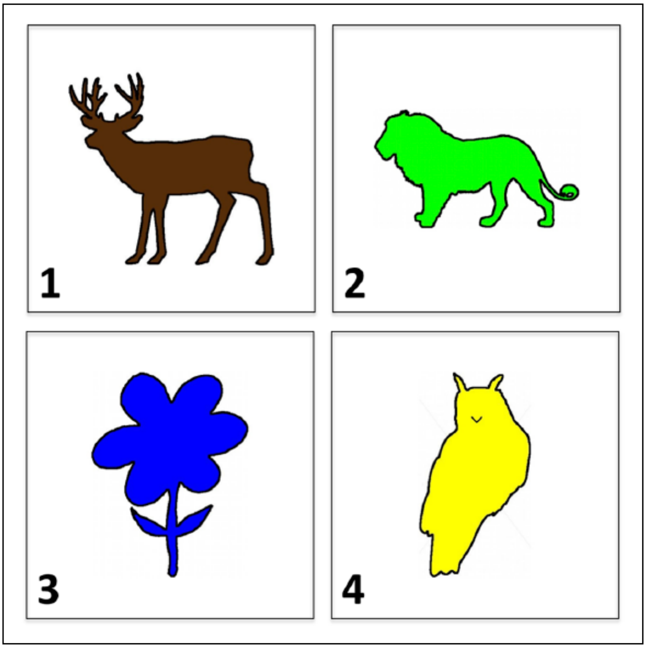

Figure 3. Example set of objects in boxes directors would see in the tile-placing task. This image depicts a brown deer in box 1, a green lion in box 2, a blue flower in box 3, and a yellow owl in box 4. This set is from a predictable NP focus trial.

The fourth block was termed the 'unpredictable' block, since in this block the global context did not fully determine which constituent was focused. This block consisted of trials with six boxes and eight available tiles. In this block, some of the available tiles differed in both shape and color, while others differed only in one of those dimensions. Therefore, the visual context was insufficient to establish which constituent (adjective, noun, or NP) would be in focus in an upcoming utterance. However, in each of these trials, the preceding linguistic context (i.e., the immediately preceding utterance) established the constituent in focus for the current utterance. For example, in a sequence of *green owl* followed by *brown owl*, the second utterance bears focus on the adjective *brown*, which is expected to be manifested phonetically. See Turnbull et al. (2015) for more details on the predictability manipulation.

Therefore, in addition to a manipulation of focus based on immediate linguistic context, like the first experiment, this experiment also involved a manipulation of how predictable the type of focus itself was. In other words, the probability of the information structure of the utterance was manipulated. This manipulation can be thought of as a *second-order* manipulation: rather than the predictability of the linguistic units themselves, it is the predictability of the types of structures the linguistic units will appear in, similar to the studies reported in Bresnan, Cueni, Nikitina, and Baayen (2007). This predictable/unpredictable manipulation of information structural probability will be referred to as 'context condition', since it relates to the overall non-linguistic context of the trial.

The final experimental manipulation that can be explored in this corpus is probability of content, another kind of utterance probability. As each trial progressed, the number of available tiles—that is, tiles visually accessible to both participants—decreased, since they were being placed on the game board by the confederate. As such, the set of possible instructions decreased with every additional box filled. The probability of any particular tile being chosen therefore increased with each subsequent instruction. Since all tiles were unique, the probability *P* of any tile being selected was 1/*N*, where *N* is the total number of available objects. It is clear that as *N* decreases, *P* increases. In this experiment, then, 'box number' was an operationalization of the probability of utterance content. The probability of an utterance at box 4 was significantly higher than the probability of the same utterance at box 1. Since the predictable/unpredictable context manipulation also involved trials of different sizes (i.e., four boxes versus six boxes), for modeling purposes the box number was normalized to a real number between 0 and 1, where 0 is the first box and 1 is the last box. This normalization allows for valid comparison of utterance probability effects between the context conditions.

An alternative analytic method could model raw probability of an individual tile, rather than box number. Box number was chosen because it was thought to reflect more closely the *predictability* of the tile, rather than its *probability*, bearing in mind the distinction between the two that was defined in the introduction. Keeping track of how many boxes remain to be filled, rather than of the reciprocal of the remaining tiles, seems to be a more intuitive measure of progress in the task. Raw probability also obscures intuitively different contexts: for example, every tile for the fourth box in a trial of six boxes (the unpredictable condition) has a probability of 0.2, which is the same probability as every tile for the first box in a trial of four boxes (the predictable condition). Using (normalized) box number abstracts away from these issues.[5]

3.3 Acoustic measurements

Burdin et al. (2015) and Turnbull et al. (2015) described the ToBI-annotation of their data by multiple trained annotators, which was found to be reliable. Their dataset provides ToBI-annotated data from 10 native speaker participants, which is analyzed here. The F0 of high and low pitch targets associated with the ToBI annotations of each word was measured in Hz, again with a frame duration of 10 ms and Boersma's (1993) autocorrelation algorithm. This measurement system means that unaccented words do not have any F0 measurements; words with H* only have F0 maxima measured; and rising accents such as L+H* have F0 minima and maxima measured. Again, duration was used as a point of comparison to contextualize the F0 results: the duration of each target word was measured in milliseconds.

As noted, minima were measured only for words where they were phonologically specified (that is, words with a L+H* accent). The relationship between F0 minima and prosodic prominence is poorly understood, partly due to a lack of clarity in the literature on what exactly "prominence" entails (see also Wagner et al., 2015), and as such the predicted behavior of F0 minima is unclear. Under a simple hypothesis where prominence is a monotonic function of F0, increasing prominence would simply be a matter of higher F0 minima. On the other hand, a more sophisticated hypothesis that holds that prominence is related to more peripheral values within a talker's pitch space would predict lower F0 minima to be more prominent. It is likewise possible that the distinction between phonologically specified minima and incidental phonetic minima is relevant to both the perception and production of prominence (see also Kochanski, Grabe, Coleman, & Rosner (2005) on the purported relationship between prominence and F0, and Goldrick, Vaughn, & Murphy (2013) on the question of what it means for a phonetic dimension to be "enhanced" or "reduced").

3.4 Data analysis

The data were analyzed using stepwise mixed-effect regression modeling. The same backwards best-path model selection algorithm as in the analysis of the first corpus was employed. As in the previous experiment, the pitch accents were collapsed into three categories of rising, high, and unaccented.

Table 8. Tables of coefficients and log-likelihood comparisons for each retained fixed effect, adjective F0 maximum model in the tile-placing experiment.

Variable	β	*SE*	*t*
Adjective: rising accent	16.751	1.651	10.146
Box number	−18.800	2.345	−8.017
Focus: adjective	−2.683	1.805	−1.486
Focus: noun	−5.212	1.952	−2.670
Context: unpredictable	−1.726	1.847	−0.935
Focus: adjective × Context: unpredictable	9.650	2.869	3.363
Focus: noun × Context: unpredictable	−4.114	3.080	−1.336
Variable	χ^2	*DF*	*p*
Adjective: rising accent	93.898	1	< .001
Box number	61.042	1	< .001
Focus condition	21.614	2	< .001
Context condition	0.027	1	.869
Focus × Context	19.549	2	< .001

Six linear models were constructed, to predict adjective F0 maximum, noun F0 maximum, adjective F0 minimum, noun F0 minimum, adjective duration, and noun duration. The fixed effects were adjective pitch accent (reference level: high), noun pitch accent (reference level: high), box number (normalized to scale from 0 to 1), context (predictable vs. unpredictable), focus (noun vs. adjective vs. NP; reference level: NP), and interactions between focus and box number, and focus and context. Random intercepts of participant, adjective, and noun identity were added. All continuous variables were centered at zero prior to being entered into the model; duration was log transformed. The two models of duration also featured the presence or absence of a prosodic break (ToBI break size of 3 or greater) after the target word as an additional fixed effect, to control for effects of phrase-final lengthening.

3.5 Results

A total of 707 target words were extracted from the corpus. As mentioned above, only the words with pitch accents had F0 peaks extracted, and only the words with rising pitch accents had F0 minima extracted. Any measurement greater than three standard deviations from the subject mean was removed. This process resulted in a removal of eight adjective F0 peak measures, three adjective F0 minimum measures, two adjective duration measures, seven noun F0 peak measures, three noun F0 minimum measures, and one noun duration measure. In sum, a total of 609 adjective F0 peak measures, 197 adjective F0 minimum measures, 705 adjective duration measures, 623 noun F0 peak measures, 317 noun F0 minimum measures, and 706 noun duration measures were analyzed.

3.5.1 Adjective F0 maximum. In the adjective F0 maximum model, four simple effects and one interaction were retained in the model. The output of this model is shown in Table 8. A significant effect of adjective pitch accent was observed, such that adjectives with a rising pitch accent had a higher peak F0 (M = 186 Hz, SD = 70.3 Hz) than adjectives with a high pitch accent (M = 168 Hz, SD = 56.0 Hz), consistent with previous research on the phonetics of these tunes in English (e.g., Bartels and Kingston, 1994; Calhoun, 2012). An effect of focus condition was observed: in the

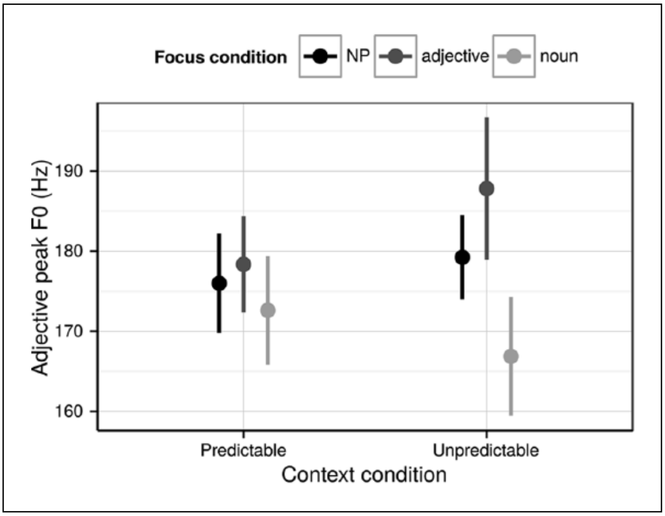

Figure 4. Means and standard errors of adjective F0 peak in different focus conditions and context conditions.

noun focus condition, the adjective F0 peak was significantly lower (*M* = 174 Hz, *SD* = 63.3 Hz) than in the NP focus condition (*M* = 177 Hz, *SD* = 64.7 Hz), suggesting that, regardless of pitch accent, the prominence of the adjective was suppressed in the noun focus condition, presumably to enhance the prominence of the noun. Although the adjective F0 peak in the adjective focus condition (*M* = 183 Hz, *SD* = 70.0 Hz) was even higher than the NP focus condition, this difference did not reach significance. However, this difference was significant when the interaction between focus and context was considered, suggesting that focus has different effects in predictable versus unpredictable contexts. Figure 4 shows means and standard errors of adjective peak F0 split by the different context and focus conditions. It can be seen that the effects of focus are largely confined to the unpredictable condition, with the effect sizes in the predictable condition remaining minimal. This result is consistent with our hypothesis: greater prosodic cues are used to enhance a contrast in an unpredictable context than in a predictable context.

Finally, a simple effect of box number was also observed, such that peak F0 fell as the box number increased. This effect is visualized in Figure 5, which shows mean adjective F0 peaks as a function of box number. Again this result is consistent with the hypothesis that utterances with a lower probability (i.e., the lower box numbers earlier in the trial) are realized with more extreme prosodic cues.

3.5.2 Noun F0 maximum. In the noun F0 maximum model, retained fixed effects included adjective pitch accent type, noun pitch accent type, focus condition, box number, and context (predictable vs. unpredictable). The model output is shown in Table 9. The peak F0 of the noun was influenced by the type of pitch accent on the preceding adjective; these differences were not significant, but trending in the expected directions. As expected, the pitch accent of the noun itself affected the peak F0, such that nouns with a rising accent had a higher peak (*M* = 178 Hz, *SD* = 66.4 Hz) than nouns with a high accent (*M* = 163 Hz, *SD* = 56.0 Hz). Likewise, focus condition had its expected effect on noun F0: nouns in the noun focus condition had higher F0 peaks (*M* = 178 Hz, *SD* = 66.0 Hz) than nouns in the NP focus condition (*M* = 169 Hz, *SD* = 58.7 Hz), which were in turn higher than noun peaks in the adjective focus condition (*M* = 166 Hz, *SD* = 58.5 Hz). These

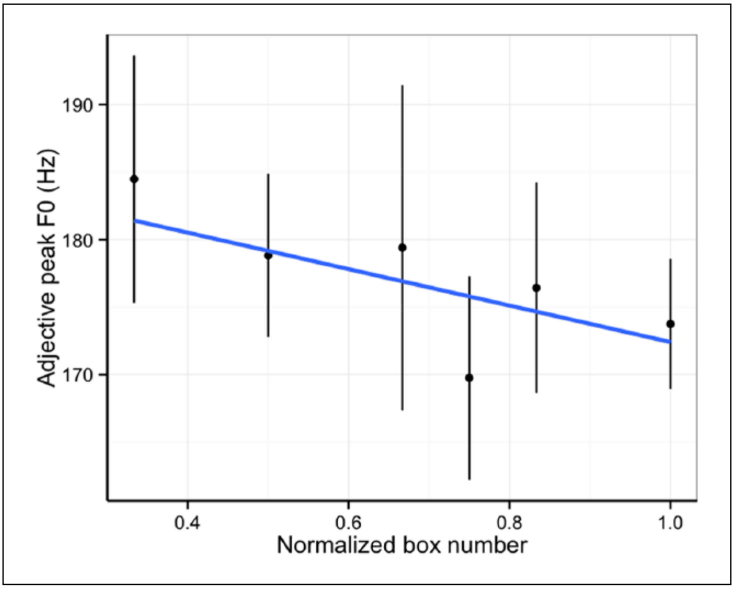

Figure 5. Relationships between (normalized) box number and mean adjective F0 peak. Bars are one standard error. Linear trend overlaid.

Table 9. Tables of coefficients and log-likelihood comparisons for each retained fixed effect, noun F0 maximum model in the tile-placing experiment.

Variable	*β*	*SE*	*t*
Adjective: no accent	2.904	2.129	1.364
Adjective: rising accent	−3.648	1.868	−1.953
Noun: rising accent	11.904	1.596	7.460
Box number	−20.653	2.620	−7.881
Focus: adjective	−5.468	1.702	−3.213
Focus: noun	3.738	1.632	2.290
Context: unpredictable	−2.603	1.362	−1.911
Variable	χ^2	*DF*	*p*
Adjective pitch accent	7.102	2	.029
Noun pitch accent	53.241	1	< .001
Box number	59.095	1	< .001
Focus condition	22.928	2	< .001
Context condition	3.640	1	.056

results are again consistent with the prominence-enhancing and prominence-suppressing roles of focus.

Overall, noun F0 peaks were lower in trials in the unpredictable context (M = 168 Hz, SD = 56.8 Hz) as compared with the predictable context (M = 173 Hz, SD = 63.6 Hz), but this difference did not reach significance. This trend is in the opposite direction to that predicted by the SSRH: higher F0 values are expected in the unpredictable context.

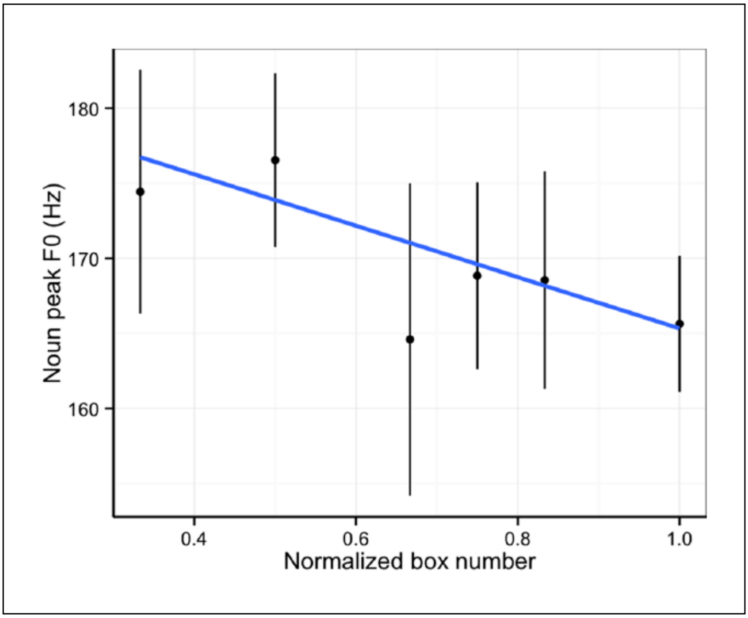

Figure 6. Relationships between (normalized) box number and mean noun F0 peak. Bars are one standard error. Linear trend overlaid.

Again, like in the adjective peak F0 model, a significant effect of box number was observed: the higher the box number, the lower the peak noun F0. See Figure 6 for a visual depiction of this effect. This effect cannot be attributed to declination, as each trial consisted of four (or six) separate utterances, with actions in between (moving the object to the appropriate location). Since each utterance was part of a separate intonational phrase, we expect pitch reset to have occurred between utterances.[6] Furthermore, this effect can also not be attributed to a general declination throughout the experiment, or participant fatigue. Although the corpus data did not include information on trial order, the experiment always had the unpredictable trials follow the predictable trials. Any declination or fatigue effect should manifest as a decrease in F0 between the predictable and unpredictable trials. Comparing mean peak F0 between the two conditions did not reveal any significant differences, either by subject, $t(9) = -0.935$, $p > 0.05$, or by item, $t(9) = -1.868$, $p > 0.05$. The only remaining plausible interpretation for this effect is then utterance predictability.

3.5.3 Adjective F0 minimum. Only two fixed effects were retained in the adjective F0 minimum model: box number and focus condition. The model output is summarized in Table 10. Adjective minima were lower in the adjective focus condition than in the NP focus condition, consistent with the hypothesis of more extreme prosodic cues (higher highs and lower lows) on material in focus compared with material not in focus.

As shown in Figure 7, an effect of box number was again observed such that lower F0 values were associated with higher box numbers (greater probability). This effect is consistent with the box number effects also observed for adjective and noun F0 peak.

3.5.4 Noun F0 minimum. No fixed effects were retained in the noun F0 minimum model. This result suggests that the noun F0 minimum was unaffected by any of the experimental manipulations.

Table 10. Tables of coefficients and log-likelihood comparisons for each retained fixed effect, adjective F0 minimum model in the tile-placing experiment.

Variable	β	*SE*	*t*
Box number	−13.708	2.257	−6.074
Focus: adjective	−3.295	1.298	−2.538
Focus: noun	2.068	1.757	1.176
Variable	χ^2	*DF*	*p*
Box number	33.619	1	< .001
Focus condition	12.740	2	.002

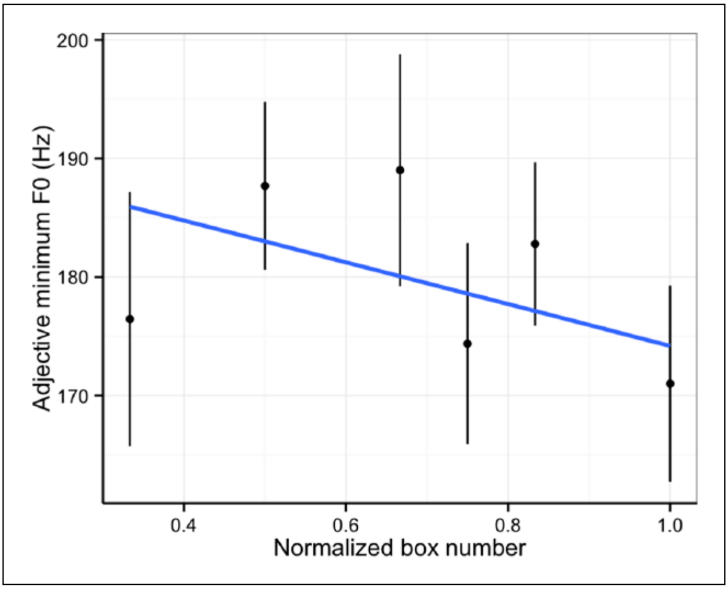

Figure 7. Relationships between (normalized) box number and mean adjective F0 minimum. Bars are one standard error. Linear trend overlaid.

3.5.5 Adjective duration. The model output for the adjective duration model is shown in Table 11. As can be seen, fixed effects of adjective pitch accent, noun pitch accent, and the presence of a prosodic break after the adjective were all retained in the model. As expected, adjectives were shorter when they were unaccented (M = 305 ms, SD = 71.6 ms) than when they had a high pitch accent (M = 331 ms, SD = 51.7 ms), and longer when they had a rising pitch accent (M = 341 ms, SD = 70.5 ms). In addition, adjectives were shorter when they were followed by an unaccented noun (M = 304 ms, SD = 56.2 ms) compared with when they were followed by a noun with a high pitch accent (M = 352 ms, SD = 71.4 ms). Consistent with phrase-final lengthening, adjectives were substantially longer when they were followed by a phrase boundary (M = 458 ms, SD = 97.4 ms) than when they were not (M = 334 ms, SD = 55.7 ms). Notably, no effects of focus or box number were observed.

3.5.6 Noun duration. In contrast with the adjective duration results, the noun duration results were considerably more complex, revealing a multiplicity of factors affecting noun duration, summarized

Table 11. Tables of coefficients and log-likelihood comparisons for each retained fixed effect, adjective duration model in the tile-placing experiment.

Variable	β	*SE*	*t*
Adjective: no accent	−73.867	19.898	−3.712
Adjective: rising accent	41.995	16.569	2.535
Noun: no accent	−73.885	22.425	−3.295
Noun: rising accent	−2.517	14.275	−0.176
Adjective is pre-pausal	247.341	32.192	7.683
Variable	χ^2	*DF*	*p*
Adjective pitch accent	26.089	2	< .001
Noun pitch accent	11.147	2	.004
Adjective is pre-pausal	56.681	1	< .001

Table 12. Tables of coefficients and log-likelihood comparisons for each retained fixed effect, noun duration model in the tile-placing experiment.

Variable	β	*SE*	*t*
Adjective: no accent	−31.176	18.860	−1.669
Adjective: rising accent	24.699	16.026	1.562
Noun: no accent	−39.538	21.295	−1.875
Noun: rising accent	32.763	14.891	2.224
Box number	−15.902	36.769	−0.436
Focus: adjective	−27.824	37.904	−0.741
Focus: noun	−34.314	37.305	−0.929
Context: unpredictable	22.019	18.517	1.201
Noun is pre-pausal	98.958	16.775	5.964
Box number × Focus: adjective	−136.500	53.614	−2.570
Box number × Focus: noun	67.245	52.707	1.289
Focus: adjective × Context: unpredictable	88.881	28.164	3.185
Focus: noun × Context: unpredictable	42.358	28.183	1.517
Variable	χ^2	*DF*	*p*
Adjective pitch accent	6.483	2	.039
Noun pitch accent	10.398	2	.006
Box number	2.166	1	.141
Focus condition	37.84	2	< .001
Context condition	27.225	1	< .001
Noun is pre-pausal	34.443	1	< .001
Box number × Focus	14.536	2	< .001
Focus × Context	10.058	2	.007

in Table 12. Fixed effects of adjective pitch accent, noun pitch accent, box number, focus condition, context condition, prosodic break after the noun, and interactions between focus and box number, and focus and context condition were all retained in the model. The inclusion of adjective pitch accent significantly improved data likelihood, but none of the individual effects were significant.

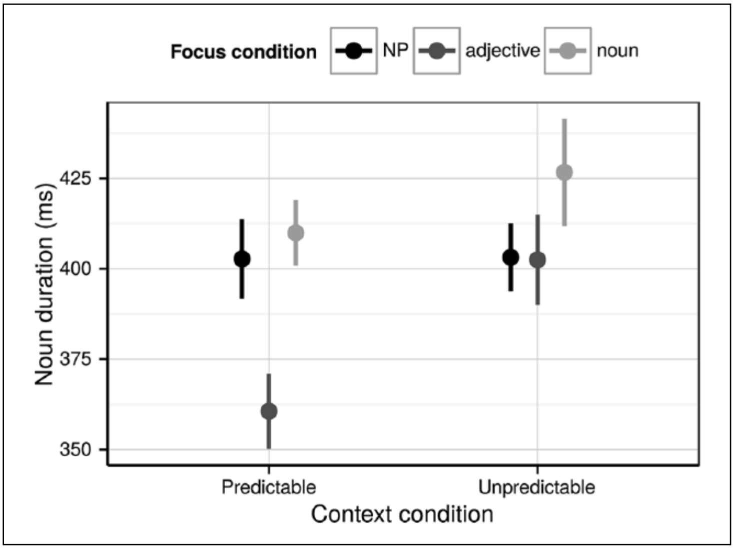

Figure 8. Relationships between focus condition, context, and noun duration.

The noun pitch accent effects were in the expected direction: nouns with rising pitch accents (*M* = 416 ms, *SD* = 94.8 ms) were longer than nouns with high pitch accents (*M* = 381 ms, *SD* = 93.5 ms), which were in turn marginally longer than unaccented nouns (*M* = 336 ms, *SD* = 57.0 ms). Again consistent with phrase-final lengthening, nouns were significantly longer when followed by a boundary (*M* = 425 ms, *SD* = 82.0 ms) than when not (*M* = 367 ms, *SD* = 130.3 ms).

Figure 8 illustrates the interaction between context condition and focus condition on noun duration. As expected, nouns in the adjective focus condition are much shorter than nouns in the NP focus condition, and nouns in the noun focus condition are much longer than nouns in the NP focus condition. The curious interaction is that the former effect (noun shortening under adjective focus) only occurs in the predictable context, while the latter effect (noun lengthening under noun focus) only occurs in the unpredictable context. To the extent that the unpredictable context represents lower overall predictability, this result is consistent with the hypothesis that unpredictable utterances have longer and more prominent words. In other words, nouns in the noun focus condition are consistently longer than nouns in the adjective focus condition, and nouns are longer in the unpredictable condition than in the predictable condition.

Both focus and box number were retained in the model, although their effects can only be understood in terms of their interaction. Figure 9 depicts this interaction. As can be seen, noun duration fell as a function of box number in both the NP focus and adjective focus conditions, consistent with the hypothesis that greater probability leads to reduced acoustic prominence. There was no effect of box number in the noun focus condition, suggesting that focus marking interferes with probability in this context.

3.6 Discussion

As observed in the Christmas tree decoration task, the results from the tile-placing task revealed expected effects of pitch accenting. Words with pitch accents were realized with higher F0 and longer duration than words without pitch accents. Again in parallel with the other experiment, the

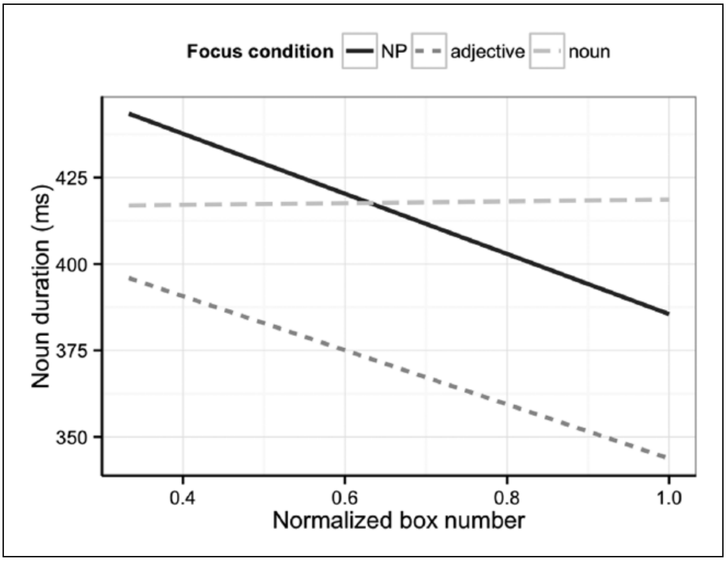

Figure 9. Noun duration as a function of box number, split by focus condition. Lines indicated linear trend.

pitch accents of both the adjective and the noun were necessary to model the F0 of either word, further underscoring the holistic nature of F0 production and the fact that pitch accents cannot be conceived of as invariant acoustic targets.

A consistent and clear effect of box number, one of our operationalizations of utterance probability, was observed for all of the F0 measures, with the exception of the noun F0 minimum. Overall, as box number increased (i.e., as the utterance probability increased), the F0 maximum and minimum of the adjective and the F0 maximum of the noun were all observed to decrease (Figures 5, 6, and 7). These effects are all consistent with the hypothesis that predictability affects F0 production such that greater, more effortful production is associated with lower-predictability linguistic material. Similarly, the noun vowel duration also fell as box number increased (Figure 9), a finding consistent with classic studies on the relationship between duration and predictability (e.g., Aylett & Turk, 2004). However, this effect was not observed in the noun focus condition, either suggesting that probability and focus interact such that focus can 'override' utterance predictability effects, or that some phrase-final lengthening process applicable to the focused nouns interfered with the predictability effects.

Context condition—predictable or unpredictable—served as the other operationalization of utterance probability in this study. This manipulation was a 'higher-order' type of predictability, relating to whether the *structure* of the upcoming utterance—specifically, which constituents would be focused—was predictable from the context or not. The effects of this manipulation were more subtle and less consistent than those of box number, but nonetheless consistent with our hypothesis of greater acoustic prominence on less predictable items. As Figure 4 shows, adjective peak F0 was more variable and more clearly affected by focus in the unpredictable condition relative to the predictable condition. This result is consistent with the notion that in the unpredictable condition, the speaker makes more use of F0 to signal which constituent is focused, while in the predictable condition the context suffices to signal the focused constituent. Figure 8 demonstrates a similar pattern for noun duration.

Also observed in these data is an asymmetry between effects on the adjective and the noun. This effect is similar to that seen in the Christmas tree decoration task. Focus condition was not found to affect adjective vowel duration; this effect appears at first to contradict previous findings that focused material is lengthened relative to material not in focus (Brown & McGlone, 1974; Eady & Cooper, 1986; Weismer & Ingrisano, 1979; Xu & Xu, 2005). However, the finding is consistent with another analysis of this dataset (Burdin et al., 2015), and could plausibly be attributed to effects of pitch accenting, which the above-cited studies did not control for. Since rising pitch accents tend to be used for material in focus, and words with rising pitch accents also tend to be longer than words with other or no pitch accents (Cambier-Langeveld & Turk, 1999; Turk & Sawusch, 1997), it stands to reason that any observed effect of focus on duration may be plausibly attributed to pitch accenting. On the other hand, an effect of focus on duration *was* observed for the nouns; likewise, Katz & Selkirk (2011) observed an effect of focus on duration independently of pitch accenting.[7] Therefore, the difference between the results of the adjective (where focus is not relevant for vowel duration) and the noun (where focus is relevant for vowel duration) must be attributed to positional or syntactic differences between the two sets of words. As mentioned in the discussion section, determining the specific locus of this effect requires careful cross-linguistic study, beyond the scope of the present paper.

4 General discussion

This paper has analyzed two corpora of speech data drawn from experiments designed to elicit prosodic contours in different contexts. These experiments, and their analyses, had several important similarities. Both experiments involved college-aged monolingual American English participants. Both tasks were interactive, with a naïve director and a confederate follower, where the director's speech was spontaneous and natural. The utterances analyzed were all imperatives involving noun phrases with 'the adjective noun' structure. The location of focus in these utterances was systematically varied as part of the experimental design, and both experiments involved some other manipulation(s) of predictability in addition to focus. The current analyses of these corpora used the same backwards stepwise linear mixed-effects regression method and controlled for similar factors.

There were also notable differences between these corpora. In terms of acoustic measures, in the Christmas tree decoration task, vowel duration was measured, as was F0 maximum on all words. In the tile-placing task, word duration was measured, and F0 maximum was only measured for words with a pitch accent. In addition, F0 minimum was measured for words with a rising pitch accent. In terms of method, the participants in the Christmas tree decoration task were given no instructions or direction about how to speak, and the speech can be considered truly spontaneous and natural. In the tile-placing task, participants were instructed to follow a set formula for all of their utterances. In the Christmas tree decoration task, only the confederate follower could see the full set of ornaments available; that information was masked from the director. In the tile-placing task, the full set of tiles was visually available to both interlocutors. This fact made the manipulation of box number as a measurement of utterance probability possible for the tile-placing task but impossible for the Christmas tree decoration task.

Similar patterns of results were obtained in both experiments. The realization of F0 was affected by the pitch accent of the word in question—rising pitch accents generally had the highest F0, followed by high pitch accents, followed by unaccented words. In both experiments, the pitch accents of both the noun and the adjective needed to be taken into consideration to accurately model F0 of either word, suggesting that, as with segmental phonemes, pitch accents should not be conceived of as invariant acoustic targets. It is well known that coarticulation is a common phenomenon in production (Öhman, 1966), and the existence of tonal coarticulation has been known for over 20 years (Xu, 1994). To a smaller extent, F0 was influenced by focus condition, such that material in focus tended to have higher F0 excursions than material not in focus.

In addition, effects of predictability were observed in both experiments, but only some types of predictability had an influence on F0. Subsequent mentions of words were subject to second mention reduction, but only in the temporal domain—F0 was not affected. The probability of utterance content, parameterized as box number in the tile-placing task, influenced F0 such that more probable utterances had lower F0 than less probable utterances. Similarly, the probability of utterance form—the 'higher-order' manipulation of predictable versus unpredictable context—interacted with focus condition to lead to more acoustic prominence on structurally unpredictable utterances, in both F0 and word duration. Notably, these findings hold regardless of pitch accent.

These effects are predicted under the hypothesis that F0 is influenced by contextual predictability. Specifically, within the theoretical framework of the SSRH, the prosodic prominence of a word is predicted to be inversely proportional to the predictability of that word in context. The results of the experiments discussed in this paper demonstrate that F0, a measure of prosodic prominence, is inversely correlated with the probability of utterance content, a measure of contextual predictability, a clear confirmation of the SSRH's prediction.

However, no effect was observed for some measures of predictability. The effects of second mention on temporal reduction and word intelligibility are well established (Baker & Bradlow, 2009; Bard & Anderson, 1994; Bard et al., 1989, 2000; Fowler & Housum, 1987; Fowler et al., 1997), and the results of this study were no exception—second and subsequent mentions of words tended to have shorter durations than the first mention. However, mention was not observed to influence F0. This lack of effect underscores the need for careful control and parameterization of 'predictability' within an experimental context, and for attentive comparison of results from different studies with different predictability manipulations. Taking into account the distinction between objective probability on the one hand and subjective predictability on the other may be a step in the right direction. This conclusion is also consistent with observations that language users are sensitive to some kinds of predictability, but not all. For example, Becker, Ketrez, and Nevins (2011) found that Turkish speakers generalize some of the language's statistical regularities to novel words; however, not all regular patterns generalize. Language users are also sensitive to the communicative context: Bosker, Quené, Sanders, and de Jong (2014) found that Dutch listeners were less likely to use disfluencies (such as *uh*) to predict upcoming material if they were listening to a non-native talker rather than a native talker. These two studies, among others, highlight the complex role of predictability in language use.

5 Conclusion

The main hypothesis of this paper has been upheld by the results of the analysis: contextual predictability has an effect on the spontaneous realization of F0 in American English. This result depends on how predictability is defined. When defined as discourse mention, predictability only affected duration, not F0. The finding of durational reduction is consistent with previous studies on second mention reduction (Fowler & Housum, 1987); the finding that F0 is unaffected is a novel one. Predictability, considered as semantic focus (Rooth, 1992), was reflected in F0 and (to a lesser extent) word and vowel duration. This finding is of course not new, although the conception of predictability as a reliable correlate of semantic focus has received little attention in the literature (with the notable exceptions of Aylett & Turk, 2004; and Calhoun, 2010b; cf. Wagner & Klassen, 2015). The predictable/unpredictable manipulation of the tile-placing experiment had inconsistent effects on F0 and duration, when considered in interaction with the focus condition of the utterance.

However, the trend of the effects was clearly toward phonetic enhancement of unpredictable items and phonetic reduction of predictable items (Turnbull et al., 2015). Finally, the clearest effect of predictability came from the 'box number' variable of the tile-placing experiment, which was a reflection of the individual utterance probability given the context. In this manipulation, an utterance was relatively improbable when there were several available objects to choose from, and the same utterance was relatively probable when there were very few available objects. An inverse correlation between utterance probability and F0 was observed, such that improbable utterances were produced with higher F0 values than probable utterances.

The types of predictability that did not yield effects of F0 are just as, if not more, revealing than the types of predictability which did yield results. The observed duration effects were generally robust and regular, in contrast with the somewhat more irregular F0 effects. A speculative functional explanation for this pattern relies on the fact that duration is not directly or primarily contrastive in English for signaling phonemic or semantic differences. A talker is therefore free to manipulate duration in order to signal predictability; put another way, duration can be modified and changed by cognitive processes under the influence of predictability without adversely interfering with the semantic content of the message. F0, on the other hand, *is* directly relevant to propositional content in English, in both intonational contrasts (such as the distinction between statements and yes–no questions, see Hadding-Koch & Studdert-Kennedy, 1964) and stressed versus unstressed syllables. Therefore, F0 cannot be as freely modified as duration without potentially interfering with semantic content.

Although all four of the predictability manipulations—focus condition, discourse mention, box number, and contextual predictability—have been presented here as potential reflections of a notion of 'predictability', it remains to be shown that they form a unified category. Indeed, the disparity of the results between each of these manipulations suggests that they each differ in their cognitive reality and subsequent effects on linguistic structure. These findings highlight the importance of carefully defining and testing any proposed metric of 'predictability' before assuming that it can be accounted for by the same mechanisms as phenomena involving other metrics (see also Watson, 2010).

These results add to the empirical support for the theoretical claims of the SSRH (Aylett & Turk, 2004). More generally, these results are broadly consistent with proposals similar to the SSRH, which suggest that phonetic material is modulated by predictability (e.g., Jaeger, 2010; Lindblom, 1990; Piantadosi, Tily, & Gibson, 2011). F0 is one of the many phonetic variables under the influence of linguistic predictability.

Acknowledgements

This work would not have been possible without the experimental data generously shared by Rachel Burdin, Cynthia Clopper, Kiwako Ito, Shari Speer, and Judith Tonhauser. I am also grateful to them for many insightful discussions of the data and experimental designs. Additionally, I would like to thank Jason Bishop, Cynthia Clopper, Sun-Ah Jun, William Schuler, Bridget Smith, Shari Speer, Kodi Weatherholtz, and an anonymous reviewer for comments on previous drafts of this paper. Any errors, of course, remain my own.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Notes

1. Hockett's observation that "in linguistics it would be an idle dream to hope for terminological agreement" remains as apt today as it was 50 years ago (Hockett, 1967, p82, fn2).

2. The NP is the largest unit considered here, since it refers to the target ornament. Therefore, depending on how focus is defined, what is referred to here as 'NP focus' could be considered equivalent to 'broad focus' or even 'no focus', since all the units within the NP are equally semantically prominent.
3. Because the target words were from a closed set, controlling for neighborhood density was not necessary (Clopper, Pisoni, & Tierney, 2006; Trout, 2005).
4. Models with random slope structures did not reliably converge, likely due to the relatively small number of data points (ranging from 780 to 842, see Results section).
5. Nevertheless, at the request of a reviewer, the modeling procedure described below was repeated with raw probability replacing box number, obtaining qualitatively similar results. This finding is not altogether surprising, since both normalized box number and raw probability are monotonically increasing throughout the course of the experiment (although the former is liner and the latter exponential). Further refinements of the probability/predictability distinction, including assessment of its necessity, is left to future work.
6. Some studies, particularly those couched in the British school's approach to intonation, have made claims about 'paragraph declination' or 'supradeclination' (e.g., Swerts & Geluykens, 1994; Wichmann, 2000), however the evidence for these effects is limited to monologues, rather than true dyadic interactions involving turn-taking and actions (such as placing tiles onto boxes) between utterances.
7. Katz & Selkirk (2011) did not directly control for pitch accenting in their modeling; however, they inferred that pitch accents were unrelated to the acoustic differences in their data after an analysis of the types of pitch accents observed between their different conditions.

References

Arnold, J. E., Kahn, J. M., & Pancani, G. C. (2012). Audience design affects acoustic reduction via production facilitation. *Psychonomic Bulletin & Review*, *19*, 505–512.

Aylett, M. (2000). *Stochastic suprasegmentals: Relationships between redundancy, prosodic structure and care of articulation in spontaneous speech*. PhD thesis, University of Edinburgh.

Aylett, M., & Turk, A. E. (2004). The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Language and Speech*, *47*(1), 31–56.

Aylett, M., & Turk, A. E. (2006). Language redundancy predicts syllabic duration and the spectral characteristics of vocalic syllable nuclei. *Journal of the Acoustical Society of America*, *119*, 3048–3058.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*, 390–412.

Baker, R., & Bradlow, A. R. (2009). Variability in word duration as a function of probability, speech style, and prosody. *Language and Speech*, *52*(4), 391–413.

Bard, E. G., & Anderson, A. H. (1994). The unintelligibility of speech to children: Effects of referent availability. *Journal of Child Language*, *10*, 265–292.

Bard, E. G., Anderson, A. H., Sotillo, C., Aylett, M., Doherty-Sneddon, G., & Newlands, A. (2000). Controlling the intelligibility of referring expressions in dialogue. *Journal of Memory and Language*, *42*, 1–22.

Bard, E. G., Lowe, A., & Altmann, G. (1989). The effects of repetition on words in recorded dictations. In J. Tubach & J. Mariani (Eds.), *Proceedings of Eurospeech '89*, volume 2, (pp. 573–576).

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*, 255–278.

Bartels, C., & Kingston, J. (1994). Salient pitch cues in the perception of contrastive focus. In P. Bosch & R. van der Sandt (Eds.), *Focus and natural language processing, Vol. 1: Intonation and syntax*. IBM Deutschland Informationssysteme GmbH Scientific Center, Institute for Logic and Linguistics, pp. 1–10.

Baumann, S., & Riester, A. (2012). Referential and lexical givenness: Semantics, prosodic, and cognitive aspects. In G. Elordieta & P. Prieto (Eds.), *Prosody and meaning*. Berlin: Mouton de Gruyter, pp. 119–161.

Becker, M., Ketrez, N., & Nevins, A. (2011). The surfeit of the stimulus: Analytic biases filter lexical statistics in Turkish laryngeal alternations. *Language*, *87*(1), 84–125.

Beckman, M. E., & Hirschberg, J. (1994). *The ToBI Annotation Conventions*, online manuscript. Retrieved from http://www.ling.ohio-state.edu/~tobi/ame_tobi/annotation_conventions.html.

Beckman, M. E., & Pierrehumbert, J. (1986). Intonational structure in Japanese and English. *Phonology Yearbook*, *3*, 255–309.

Bell, A., Brenier, J. M., Gregory, M., Girand, C., & Jurafsky, D. (2009). Predictability effects on durations of content and function words in conversational English. *Journal of Memory and Language*, *60*(1), 92–111.

Bell, A., Jurafsky, D., Fosler-Lussier, E., Girand, C., & Gregory, M. (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.

Birner, B. J. (1994). Information status and word order: an analysis of English inversion. *Language*, *70*(2), 233–259.

Boersma, P. (1993). Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. *Proceedings of the Institute of Phonetic Sciences*, *17*, 97–110.

Bolinger, D. (1972). Accent is predictable (if you're a mind-reader). *Language*, *48*, 633–644.

Bosker, H. R., Quené, H., Sanders, T., & de Jong, N. H. (2014). Native 'um's elicit prediction of low-frequency referents, but non-native 'um's do not. *Journal of Memory and Language*, *75*, 104–116.

Bresnan, J., Cueni, A., Nikitina, T., & Baayen, R. H. (2007). Predicting the dative alternation. In G. Bouma, I. Krämer, & J. Zwarts (Eds.), *Cognitive foundations of interpretation*, (pp.69–94). Royal Netherlands Academy of Science, Amsterdam.

Brown, W. S., & McGlone, R. E. (1974). Aerodynamic and acoustic study of stress in sentence productions. *Journal of the Acoustical Society of America*, *56*, 971–974.

Burdin, R. S., & Clopper, C. G. (2015). Phonetic reduction, vowel duration, and prosodic structure. *Proceedings of the 18th International Congress of Phonetic Sciences*.

Burdin, R. S., Phillips-Bourass, S., Turnbull, R., Yasavul, M., Clopper, C. G., & Tonhauser, J. (2015). Variation in the prosody of focus in head- and edge-marking languages. *Lingua*, *165*(B), 254–276.

Calhoun, S. (2004). Phonetic dimensions of intonational categories - The case of L+H* and H*. *Proceedings of Speech Prosody 2004*, pp.103–106.

Calhoun, S. (2010a). The centrality of metrical structure in signaling information structure: A probabilistic perspective. *Language*, *86*(1), 1–42.

Calhoun, S. (2010b). How does informativeness affect prosodic prominence? *Language and Cognitive Processes*, *25*, 1099–1140.

Calhoun, S. (2012). The theme/rheme distinction: Accent type or relative prominence? *Journal of Phonetics*, *40*, 329–349.

Cambier-Langeveld, T., & Turk, A. E. (1999). A cross-linguistic study of accentual lengthening: Dutch vs English. *Journal of Phonetics*, *27*(3), 255–280.

Chafe, W. (1974). Language and consciousness. *Language*, *50*, 111–133.

Clopper, C. G., Pisoni, D. B., & Tierney, A. T. (2006). Effects of open-set and closed-set task demands on spoken word recognition. *Journal of the American Academy of Audiology*, *17*(5), 331–349.

Cole, J., Mo, Y., & Hasegawa-Johnson, M. (2010). Signal-based and expectation-based factors in the perception of prosodic prominence. *Laboratory Phonology*, *1*, 425–452.

Cooper, W. E., Eady, S. J., & Mueller, P. R. (1985). Acoustical aspects of contrastive stress in question–answer contexts. *Journal of the Acoustical Society of America*, *77*(6), 2142–2156.

Dahan, D., Tanenhaus, M. K., & Chambers, C. G. (2002). Accent and reference resolution in spoken-language comprehension. *Journal of Memory and Language*, *47*, 292–314.

Eady, S. J., & Cooper, W. E. (1986). Speech intonation and focus location in matched statements and questions. *Journal of the Acoustical Society of America*, *80*, 402–415.

Eady, S., Cooper, W., Klouda, G., Mueller, P., & Lotts, D. (1986). Acoustical characteristics of sentential focus: Narrow vs. broad and single vs. dual focus environments. *Language and Speech*, *29*, 233–251.

Face, T. (2001). Focus and early peak alignment in Spanish intonation. *Probus*, *13*(2), 223–246.

Féry, C., & Kügler, F. (2008). Pitch accent scaling on given, new and focused constituents in German. *Journal of Phonetics*, *36*, 680–703.

Féry, C., & Samek-Lodovici, V. (2006). Focus projection and prosodic prominence in nested foci. *Language*, *82*(1), 131–150.

Fowler, C. A., & Housum, J. (1987). Talkers' signaling of "new" and "old" words in speech and listeners' perception and use of the distinction. *Journal of Memory and Language*, *26*, 489–504.

Fowler, C. A., Levy, E. T., & Brown, J. M. (1997). Reductions of spoken words in certain discourse contexts. *Journal of Memory and Language*, *37*, 24–40.

Fox Tree, J. E., & Clark, H. H. (1997). Pronouncing "the" as "thee" to signal problems in speaking. *Cognition*, *62*, 151–167.

Gahl, S. (2008). *Time* and *thyme* are not homophones: The effect of lemma frequency on word durations in spontaneous speech. *Language*, *84*(3), 474–496.

Galati, A., & Brennan, S. E. (2010). Attenuating information in spoken communication: For the speaker, or for the addressee? *Journal of Memory and Language*, *62*, 35–51.

Goldrick, M., Vaughn, C., & Murphy, A. (2013). The effects of lexical neighbors on stop consonant articulation. *Journal of the Acoustical Society of America*, *134*(2), EL172–EL177.

Grabe, E. (1998). Pitch accent realization in English and German. *Journal of Phonetics*, *26*(2), 129–143.

Grosz, B. J., & Sidner, C. L. (1986). Attention, intentions, and the structure of discourse. *Computational Linguistics*, *12*(3), 175–204.

Gundel, J. K., Hedberg, N., & Zacharski, R. (1993). Cognitive status and the form of referring expressions in discourse. *Language*, *69*(2), 274–307.

Hadding-Koch, K., & Studdert-Kennedy, M. (1964). An experimental study of some intonation contours. *Phonetica*, *11*, 175–185.

Heller, J., & Goldrick, M. (2014). Grammatical constraints on phonological encoding in speech production. *Psychonomic Bulletin & Review*, *21*(6), 1576–1582.

Hockett, C. (1967). *Language, Mathematics, and Linguistics*. Mouton & Co, The Hague.

Ito, K., & Speer, S. R. (2006). Using interactive tasks to elicit natural dialogue. In S. Sudhoff, D. Lenertová, R. Meyer, S. Pappert, P. Augurzky, I. Mleink, N. Richter, & J Schließer (Eds.), *Methods in Empirical Prosody Research* (pp.227–257). Walter de Gruyter, Berlin.

Ito, K., Speer, S. R., & Beckman, M. E. (2004). Informational status and pitch accent distributions in spontaneous dialogues in English. In *Proceedings of Speech Prosody 2004* (pp.279–282). *Nara, Japan*.

Jaeger, T. F. (2010). Redundancy and reduction: Speakers manage syntactic information density. *Cognitive Psychology*, *61*, 23–62.

Jurafsky, D., Bell, A., Gregory, M., & Raymond, W. (2001). Probabilistic relations between words: Evidence from reduction in lexical production. In J. Bybee & P. Hopper (Eds.), *Frequency and the emergence of linguistic structure* (pp.229–254). John Benjamins, Amsterdam.

Kahn, J., & Arnold, J. E. (2012). A processing-centered look at the contribution of givenness to durational reduction. *Journal of Memory and Language*, *67*, 311–325.

Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology*, *3*, 430–454.

Kahneman, D., & Tversky, A. (1982). On the study of statistical intuitions. *Cognition*, *11*, 123–141.

Katz, J., & Selkirk, E. O. (2011). Contrastive focus vs. discourse-new: Evidence from phonetic prominence in English. *Language*, *87*(4), 771–816.

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

Kohler, K. (2008). The perception of prominence patterns. *Phonetica*, *65*(4), 257–269.

Krahmer, E., & Swerts, M. (2001). On the alleged existence of contrastive accents. *Speech Communication*, *34*, 391–405.

Ladd, D. R. (2008). *Intonational Phonology*. Cambridge University Press, Cambridge, 2nd edition.

Ladd, D. R., & Schepman, A. (2003). "Sagging transitions" between high pitch accents in English: experimental evidence. *Journal of Phonetics*, *31*, 81–112.

Lam, T. Q., & Watson, D. G. (2014) Repetition reduction: Lexical repetition in the absence of referent repetition. *Journal of Experimental Psychology: Learning, Memory & Cognition*, *40*, 829–843.

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

Lindblom, B. (1990). Explaining phonetic variation: a sketch of the H&H theory. In W. J. Hardcastle & A. Marchal (Eds.), *Speech production and speech modelling* (pp.403–439). Kluwer, Dordrecht.

Lyon, D., & Slovic, P. (1976) Dominance of accuracy information and neglect of base rates in probability estimation. *Acta Psychologica*, *40*, 287–298.

Öhman, S. E. G. (1966). Coarticulation in VCV utterances: Spectrographic measurements. *Journal of the Acoustical Society of America*, *39*, 151–168.

Piantadosi, S. T., Tily, H., & Gibson, E. (2011). Word lengths are optimized for efficient communication. *Proceedings of the National Academy of Sciences*, 108(9), 3526.

Pierrehumbert, J. (1980). *The phonology and phonetics of English intonation*. PhD thesis, MIT.

Pierrehumbert, J. (2003). Probabilistic phonology: Discrimination and robustness. In R. Bod, J. Hay, & S. Jannedy (Eds.), *Probabilistic Linguistics* (pp.177–228). MIT Press, Cambridge, MA.

Pierrehumbert, J., & Hirschberg, J. (1990). The meaning of intonational contours in the interpretation of discourse. In P. Cohen, J. Morgan, & M. M. Pollack (Eds.), *Intentions in Communication* (pp.271–311). MIT Press, Cambridge, MA.

Pierrehumbert, J., & Steele, S. (1990). Categories of tonal alignment in English. *Phonetica*, *47*, 181–196.

Pluymaekers, M., Ernestus, M., & Baayen, R. H. (2005). Articulatory planning is continuous and sensitive to informational redundancy. *Phonetica*, *62*, 146–159.

Rahman, A., & Ng, V. (2011). Learning the information status of noun phrases in spoken dialogues. In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, pp. 1069–1080.

Riester, A., & Piontek, J. (2015) Anarchy in the NP. When new nouns get deaccented and given nouns don't. *Lingua*, *165*(B), 230–253.

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

Schwarzschild, R. (1999). GIVENness, AvoidF and other constraints on the placement of accent. *Natural Language Semantics*, *7*, 141–177.

Seyfarth, S. (2014). Word informativity influences acoustic duration: Effects of contextual predictability on lexical representation. *Cognition*, *133*, 140–155.

Silverman, K., Beckman, M. E., Pitrelli, J., Ostendorf, M., Wightman, C., Price, P., Pierrehumbert, J., & Hirschberg, J. (1992). ToBI: A standard for labeling English prosody. In *Proceedings of the 1992 International Conference on Spoken Language Processing*, pp. 867–870.

Swerts, M., & Geluykens, R. (1994). Prosody as a marker of information flow in spoken discourse. *Language and Speech*, *37*(1), 21–43.

Swerts, M., Krahmer, E., & Avesani, C. (2002). Prosodic marking of information status in Dutch and Italian: A comparative analysis. *Journal of Phonetics*, *30*, 629–654.

Terken, J., & Hirschberg, J. (1994). Deaccentuation of words representing 'given' information: Effects of persistence of grammatical function and surface position. *Language and Speech*, *37*(2), 125–145.

Tily, H., Gahl, S., Arnon, I., Snider, N., Kothari, A., & Bresnan, J. (2009). Syntactic probabilities affect pronunciation in spontaneous speech. *Language and Cognition*, *1*(2), 147–165.

Tily, H., & Kuperman, V. (2012). Rational phonological lengthening in spoken Dutch. *Journal of the Acoustical Society of America*, *132*(6), 3935–3940.

Trout, J. D. (2005). Lexical boosting of noise-band speech in open- and closed-set formats. *Speech Communication*, *47*, 424–435.

Turk, A. (2010). Does prosodic constituency signal relative predictability? A smooth signal redundancy hypothesis. *Laboratory Phonology*, *1*(2), 227–262.

Turk, A. E., & Sawusch, J. R. (1997). The domain of accentual lengthening in American English. *Journal of Phonetics*, *25*(1), 25–41.

Turnbull, R., Burdin, R. S., Clopper, C. G., & Tonhauser, J. (2015). Contextual information and the prosodic realization of focus: A cross-linguistic comparison. *Language, Cognition, and Neuroscience*, *30*(9), 1061–1076.

Turnbull, R., Royer, A. J., Ito, K., & Speer, S. R. (2014). Prominence perception in and out of context. In *Proceedings of the 7th international conference on Speech Prosody*, pp. 1164–1168.

Wagner, M., & Klassen, J. (2015). Accessibility is no alternative to alternatives. *Language, Cognition and Neuroscience*, *30*(1-2), 212–233.
Wagner, P., Origlia, A., Avesani, C., Christodoulides, G., Cutugno, F., D'Imperio, M., … Vainio, M. (2015). Different parts of the same elephant: A roadmap to disentangle and connect different perspectives on prosodic prominence. *Proceedings of ICPhS XVIII*.
Watson, D. G. (2010). The many roads to prominence: Understanding emphasis in conversation. *Psychology of Learning and Motivation*, *52*, 163–183.
Watson, D. G., Arnold, J. E., & Tanenhaus, M. K. (2008). Tic tac toe: Effects of predictability and importance on acoustic prominence in language production. *Cognition*, *106*, 1548–1557.
Weismer, G., & Ingrisano, D. (1979). Phrase-level timing patterns in English: Effects of emphatic stress location and speaking rate. *Journal of Speech and Hearing Research*, *22*, 516–533.
Wichmann, A. (2000). *Intonation in Text and Discourse: Beginnings, Middles and Ends*. Longman, Harlow, UK.
Xu, Y. (1994). Production and perception of coarticulated tones. *Journal of the Acoustical Society of America*, *95*(4), 2240–2253.
Xu, Y. (1999). Effects of tone and focus on the formation and alignment of F0 contours. *Journal of Phonetics*, *27*, 55–105.
Xu, Y., & Xu, C. X. (2005). Phonetic realization of focus in English declarative intonation. *Journal of Phonetics*, *33*, 159–197.