This is an Accepted Manuscript of a chapter in Individual Differences in Speech Production and Perception, published in 2015 by Peter Lang International Academic Publishers.

Individual differences in the prosodic encoding of informativity

Iris Chuoying Ouyang and Elsi Kaiser

Abstract

This chapter presents a psycholinguistic production study that investigates individual differences in the prosodic encoding of informativity. In particular, it examines how the shapes of f0 contours and the sizes/ranges of f0 excursions are influenced by the interaction between information structure and information-theoretic properties. We focus on two types of information structure, namely new-information narrow focus and corrective narrow focus, and two kinds of information-theoretic properties, namely word frequency and contextual probability. We analyze (i) group trends, (ii) between-subject variability as well as (iii) within-subject variability, and thereby identify speaker-specific effects. Our results show that word frequency and contextual probability modulate the f0 movement associated with new-information narrow focus and corrective narrow focus respectively (see also Ouyang & Kaiser 2014). Furthermore, f0 ranges appear to be more informative than f0 shapes in reflecting informativity across speakers. Specifically, speakers seem to have individual 'preferences' regarding f0 shapes, the f0 ranges they use for an utterance, and the magnitude of differences in f0 ranges by which they mark information-structural distinctions. In contrast, there is more universality over the *directions of differences in f0 ranges between information-structural types*. Our findings highlight the importance of disentangling information structure and information-theoretic factors and examining both inter- and intra-speaker variability.

1. Introduction

It is widely accepted that prosody can reflect the extent to which a linguistic element is 'informative'. Prior work has approached the relationship between prosody and informativity from various angles, of which two popular ones are **information structure** (e.g. Breen, Fedorenko, Wagner, & Gibson, 2010； Brown, 1983; Cooper, Eady, & Mueller, 1985; Couper-Kuhlen, 1984; Eady & Cooper, 1986; Hay, Warren, & Drager, 2006; Katz & Selkirk, 2011; Krahmer & Swerts, 2001; Ladd, 1996; Pierrehumbert & Hirschberg, 1990) and **information theory** (e.g. Aylett & Turk, 2004; Baker and Bradlow, 2009; Bell, Jurafsky, Fosler-Lussier, Girand, Gregory, & Gildea, 2003; Bell, Brenier, Gregory, Girand, & Jurafsky, 2009; Calhoun, 2010; Clopper & Pierrehumbert, 2008; Gregory, Raymond, Bell, Fosler-Lussier, & Jurafsky, 1999; Lieberman, 1963; Munson & Soloman, 2004; Pan & Hirschberg, 2000; Pitrelli, 2004; Pluymaekers, Ernestus, & Baayen, 2005a, 2005b; Scarborough, 2010; van Son, Koopmans-van Beinum, & Pols, 1999; Wright, 2003). It has been found that the acoustic properties of an utterance such as duration, f0, intensity, and spectral characteristics provide cues for the relative informativity of its components (see Wagner & Watson, 2010, for a review). However, existing studies have also noted that speakers differ in their acoustic characteristics and the prosodic

patterns they use to signal linguistic categories (e.g. Allen, Miller & Desteno, 2003; Dahan & Bernard, 1996; Fougeron, 2004; Fougeron & Kewley-Port, 2007; Loakes & McDougall, 2010; Niebuhr, D'Imperio, Fivela, & Cangemi, 2011; Smith & Hawkins, 2012; Theodore, Miller, & DeSteno, 2007; Trouvain & Grice, 1999). In this section, we will first discuss the previous research on prosody from the perspectives of information structure, information theory, and individual differences. Then, we will describe the aims and predictions of our study, which integrates the insights from these different traditions of research and furthers our understanding of prosody and informativity.

1.1. Prosodic prominence and information structure

In the information-structure-based tradition, acoustic prominence is associated with linguistic material in the foreground, or in focus – broadly speaking, material that adds new information to the conversation. Depending on the preceding discourse, speakers may emphasize particular words in an utterance to direct their addressee's attention to the important message they are trying to convey. It has been found that some types of information structure differ acoustically from each other. For instance, consider the word 'toys' in the following contexts:

(1) a. What did David find on the stairs?
    b. He found toys on the stairs. ['toys' = narrow, new-information focus]
(2) a. Did David find toys on the stairs?
    b. Yes, he found toys on the stairs. ['toys' = given information, unfocused]
(3) a. What happened?
    b. David found toys on the stairs. ['toys' = wide, new-information focus]

In response to (1a), *toys* in (1b) is in new-information focus, as it conveys information that has not been mentioned and cannot be inferred from the preceding discourse. In contrast, the same word *toys* in (2b) responding to (2a) is unfocused, given information, because what it conveys has been expressed in the preceding discourse (e.g. Prince, 1992; Rooth, 1992). Furthermore, *toys* in (1b) is narrowly focused new information, since it is the only component of the utterance that introduces new information to the conversation. However, the same word *toys* in (3b) in response to (3a) is new information in wide focus, because the entire utterance with multiple components including toys is in new-information focus (e.g. Gussenhoven, 1983; Selkirk, 1984). It has been shown that new elements are acoustically more prominent than given elements (e.g. Brown, 1983; Eady & Cooper, 1986; Hay et al., 2006; Krahmer & Swerts, 2001; Ladd, 1996), and that material in narrow new-information focus is acoustically more prominent than the same material in wide new-information focus (e.g. Breen et al., 2010; Eady & Cooper, 1986).

Another kind of information structure that has been extensively studied is contrastive focus, of which various subtypes have been identified (e.g. Vallduví & Vilkuna, 1998). For example, two

common types of contrastive focus, both involving explicit alternatives in the preceding discourse, are shown in (4-5). *Toys* in (4b) responding to (4a) picks 'toys' from the set consisting of 'books' and 'toys' that has been established via (4a), and *toys* in (5b) responding to (5a) is intended to contradicts the information *socks* that has been conveyed via (5a). In this study, we concentrate on the latter type of contrastive focus, which has been referred to as corrective focus (e.g. Dik, 1997). We chose this subtype because its information-structural properties are well-understood and it is prevalent in communication. Contrastive/corrective elements have been shown to receive greater acoustic prominence than non-contrastive/non-corrective elements, whether they are given or new material in the discourse (e.g. Breen et al, 2010; Cooper et al, 1985; Couper-Kuhlen, 1984; Katz & Selkirk, 2011; Krahmer & Swerts, 2001).

(4) a. Did David find books or toys on the stairs?
    b. He found toys on the stairs. ['toys' = narrow, contrastive focus]
(5) a. Did David find socks on the stairs?
    b. No, he found toys on the stairs. ['toys' = narrow, contrastive/corrective focus]

Various acoustic properties have been found to reflect information-structural salience, including types or presence of accents on and after the focused element (e.g. Krahmer & Swerts, 2001; Ladd, 1996; Pierrehumbert & Hirschberg, 1990), expanded vowel space and increased formant movement in the focused element (e.g. Hay et al., 2006), increased duration, f0[1], intensity, and more f0 protrusion during the focused element, more f0 compression following the focused element (e.g. Breen et al., 2010; Brown, 1983; Couper-Kuhlen, 1984; Katz & Selkirk, 2011), decreased duration, f0 and intensity preceding the focused element (e.g. Eady & Cooper, 1986), and a sudden drop or sharper fall within or following the focused element (e.g. Cooper et al, 1985; Couper-Kuhlen, 1984; Eady & Cooper, 1986).

1.2. Prosodic prominence and information-theoretic factors

In addition to work from the information-structural perspective, there is also research in the information-theoretic tradition, where a correlation has been found between acoustic reduction and the redundancy, or the predictability of a linguistic element. Depending on what is more (or less) common in the language or the given linguistic environment, certain elements may be pronounced with more or less acoustic prominence. A wide variety of probabilistic measurements have been used to represent the predictability of a segment, phoneme, or word. Examples include context-independent properties such as frequency and neighborhood density (e.g. Munson & Soloman, 2004; Pitrelli, 2004; Scarborough, 2010; Wright, 2003) and context-dependent properties such as joint probability, conditional probability, mutual information, and

[1] In Breen et al. (2010), focus breadth (narrow vs. wide) and contrastiveness (corrective vs. non-corrective) have opposite effects on f0. Narrow focus is marked with *higher* mean and maximum f0 than wide focus, while correctively focused word is produced with *lower* mean and maximum f0 than non-correctively focused word. This finding about contrastiveness diverges from other previous research.

semantic predictability (e.g. Bell et al., 2003; Clopper & Pierrehumbert, 2008; Lieberman, 1963; Pan & Hirschberg, 2000; Scarborough , 2010; van Son et al, 1999). Elements that occur more frequently or have more neighbors (i.e. items that are similar to each other due to overlapping features) in the language are acoustically more reduced than elements that occur less frequently or have fewer neighbors. Likewise, elements that are more likely to occur in a particular environment (based on adjacent items or semantic context) receive larger acoustic reduction than elements that are less likely to occur in the environment. Research has found information-theoretic predictability being realized with decreased duration and amplitude (e.g. Bell et al., 2003; Lieberman, 1963), lower likelihood of accentuation (e.g. Pan & Hirschberg, 2000; Pitrelli, 2004), lower center of gravity of the power spectrum (CoG), less extreme distance between the first and second formants (e.g. van Son et al., 1999), shorter vowels, and less dispersed vowel space (e.g. Clopper & Pierrehumbert, 2008; Munson and Soloman, 2004; Scarborough , 2010; Wright, 2003).

1.3. Connections between information-structural and information-theoretic approaches

While the information-structural and the information-theoretic traditions focus on different factors of informativity from distinct perspectives, they have found similar prosodic patterns that signal the relative degree of informativity between linguistic elements (see Sections 1.1-1.2). A higher degree of informativity in general results in more exhaustive use of a prosodic space, whichever acoustic dimension it is that a particular study examines. This leads us to the question of how information structure and information-theoretic properties interact in influencing prosody: Do they have simply have additive effects, or they interact in a non-additive way? To our knowledge, only a limited number of studies have investigated both of these two types of informativity (e.g. Aylett and Turk, 2004; Baker and Bradlow, 2009; Bell et al., 2009; Calhoun, 2010; Gregory et al., 1999; Pluymaekers et al., 2005a, 2005b). Most of these studies take an information-theoretic approach that includes the repeated use of words as a redundancy factor. Repeated words are by definition given, or at least not entirely new, information, and thus the information-theoretic notion of repetition can be regarded as givenness in an information-structural view (e.g. Fowler & Housum, 1987). The effect of word repetition, over and above (other) information-theoretic factors, has been found on different kinds of linguistic units. Aylett and Turk (2004) measure how many times a referent has been previously mentioned, and show that syllable duration decreases as the order of mention increases, in addition to the effects of word frequency and syllable conditional trigram probability. For suffixed words in Dutch, Pluymaekers, Ernestus, and Baayen (2005b) measure how many times a word has been uttered, and show that repetition significantly reduces the duration of suffixes and marginally reduces the duration of stems and entire words, in addition to the effects of mutual information with the adjacent words. Bell, Brenier, Gregory, Girand and Jurafsky (2009) find that, in English, content words are shorter when repeated, more frequent, or more predictable from the following word, while function words are not so affected by repetition and word frequency, but are affected by

the predictability from the following word. The predictability from the preceding word only shortens very frequent function words. Lastly, Gregory, Raymond, Bell, Fosler-Lussier and Jurafsky (1999) find that word duration decreases as the following redundancy factors increase: word frequency, mutual information, conditional bigram probability, semantic relatedness, and repetition. In sum, word repetition has been shown to cause shortening at the syllable, morpheme and word levels, even when we take into account word frequency and other statistical-probabilistic factors based on adjacent items or semantic context.

To the best of our knowledge, there is only one existing study that addresses the *interaction* between word repetition and (other) information-theoretic factors. In a production experiment where participants read a number of paragraphs twice, Baker and Bradlow (2009) find that word frequency influences the amount of reduction a word undergoes when it is mentioned for the second time. Higher-frequency words exhibit more shortening upon second mention than lower-frequency words, when word length is controlled. Furthermore, this interaction is only found in plain speech, i.e., when participants are instructed to speak as if they are talking to someone familiar with their voice and speech patterns. It does not occur in clear speech, i.e., when participants are instructed to speak as if they are talking to a listener with a hearing loss or to a non-native speaker learning their language. From the perspective of information structure, this finding can be restated as: the duration cue for new information (i.e. first mention) is weaker in lower-frequency words, and weaker in clear speech compared to plain speech. Thus, there seems to be a saturation effect such that the prosodic cues for information structure are weakened when information-theoretic factors also demand prosodic prominence. However, it remains unclear whether other kinds of information-theoretic factors, such as contextual probability, have a similar impact and whether other kinds of information structure, such as corrective focus, are affected in a similar way. Calhoun (2010) shows that whether a word carries a nuclear accent, non-nuclear accent, or no accent can be predicted using models including word frequency, bigram probability, the presence/absence of focus, as well as other factors. Nevertheless, no interaction between these factors is mentioned. In sum, it is not yet well-understood how information-theoretic properties and information structure interact to influence prosody.

1.4. Individual differences in prosody and the prosodic encoding of informativity

In addition to the interaction between different types of informativity, another important factor that influences an utterance's prosodic representation is individual differences. Research has shown that speakers should not be assumed homogenous even though they speak the same language. Speakers can differ in their ways of marking the linguistic distinction in question using duration, f0, intensity and spectral parameters. To name a few, individual differences have been investigated in the duration and spectral cues for word boundary (e.g. Smith & Hawkins, 2012), in voice-onset-time (VOT) for stop consonants (e.g. Allen, Miller & Desteno, 2003; Loakes &

McDougall, 2010), and how VOT is affected by other factors such as speech rate and place of articulation (e.g. Theodore, Miller & DeSteno, 2007).

It appears that between-subject variability can occur qualitatively and quantitatively, both on a general level and in specific cases. Along a given acoustic dimension, participants have different ranges of absolute values, produce different sizes and directions of variation between and within linguistic categories, and use different kinds and numbers of strategies to signal a linguistic contrast. For example, in a study where participants were asked to speak at self-selected fast, normal and slow rates, some people's fast rates were similar to some others' slow rates in terms of the number of syllables they produced per second. Moreover, the participants differed in how they altered their speech rate: while some people produced more syllables a second for a faster rate, some others produced longer pauses for a slower rate (German: Trouvain & Grice, 1999). In a study by Dahan and Bernard (1996) on French emphatic accent with four participants, some people increased f0 to a greater extent than others. The participants also differed in where and how they used intensity to signal emphasis. For the emphasized element in a sentence, one person increased the intensity, another person decreased it, and two other people produced no difference. In the sentence region preceding the emphasized element, three people decreased the intensity, while one person produced no differences. Lastly, everyone decreased the intensity in the sentence region following the emphasized element (Dahan & Bernard, 1996).

In addition to individual differences in the modulation of duration, pauses, f0 and intensity, work by Niebuhr, D'Imperio, Fivela and Cangemi (2011) found evidence for individual differences on the realization of pitch accent categories in Standard Northern German (H* vs. H+L*), Neapolitan Italian (L+H* vs. L*+H) and Pisa Italian (H* vs. H*+L). Niebuhr et al. found that Standard Northern German and Neapolitan Italian speakers used different strategies in terms of the alignment and shapes of f0 contours: some people produced systematic differences in the location of f0 peak with respect to the target syllable, while others produced systematic differences in how steep and large the f0 rise or fall was. In contrast, Pisa Italian speakers only differed in cue strength: those who made greater alignment differences also made greater differences in shapes (Niebuhr, D'Imperio, Fivela & Cangemi, 2011).

Individual differences also exist in the strategies people use for increasing the audibility/intelligibility of their speech. In a study where participants were first asked to speak normally and then asked to speak as they would if they were talking to a hearing-impaired person, individual differences were observed. According to normal-hearing listeners in a perception study, some of the speakers significantly improved their vowel intelligibility while others did not. It turns out that the former group of speakers increased their vowel duration and raised their F2 for front vowels to a greater extent than the latter group. Also, the former group expanded their vowel space in the F1 dimension, while the latter group did not (Fougeron, 2004; Fougeron & Kewley-Port, 2007). In sum, empirical evidence suggests that speakers may differ from one

another substantially in terms of whether and how particular acoustic markers correlate with particular linguistic factors.

In addition to the studies that explicitly focus on individual differences, research whose primary focus is not on individual differences has also led to observations about between-subject variability, i.e. how individuals differ. For example, it has been noted that participants differed in their duration and spectral cues for the edges of prosodic domains (e.g. Fougeron & Keating, 1997; Krivokapic & Byrd, 2012; Korean: Cho & Keating, 2001), in their pausing and lengthening cues for levels of discourse structure (e.g. word vs. clause vs. paragraph in Dutch: van Donzel & Beinum, 1996), and in the effect of word prosodic structure on vowel duration (e.g. Dutch: Rietveld, Kerkhoff & Gussenhoven, 2004).

More specifically related to informativity, Krahmer and Swerts (2001) investigated the intonational cues for the distinctions between contrastive focus, non-contrastive focus, and given information in Dutch. An interactive task was used, where participants worked in pairs to complete dialogues. It was found that some participants' prosodic behavior ignored their partner's contribution and instead prosodically marked elements that were contrastive to their own last utterance. These participants also tended to end their utterances with a high boundary tone (H%), which is generally interpreted as signaling the speaker's intention to hold the turn. Thereby, these 'egocentric' participants made the exceptional cases in the data.

In related work on focus types, Andreeva, Barry and Steiner (2007) investigated the cues in duration, f0, intensity and vowel quality for the distinctions between narrow contrastive focus, narrow non-contrastive focus, and wide focus in German. They note that some participants produced larger differences than others, and some participants also used one parameter to a greater extent than another. Thus, individual participants had their own tendencies and strategies in producing prosodic prominence. Other than these sparse observations, little is known about the extent or nature of individual differences regarding the prosodic encoding of informativity.

1.5. The present study: Aims and expected outcome

The previous research discussed in Sections 1.1-1.3 shows that an utterance's prosodic representation depends on how informative each of its constituents is. Information-structural status, such as being in narrow focus, and information-theoretic properties, such as lexical frequency and contextual probability, both play a role in prosody. It is striking that little attention has been paid to the potential interaction between information structure and information-theoretic factors, given the considerable efforts that have been devoted to both kinds of factors separately. To shed light on this issue, we conducted a psycholinguistic production study (see Ouyang & Kaiser 2014 for an earlier discussion of this study) to investigate whether information structure and information-theoretic factors interact in determining a word's prosodic prominence,

and if so, whether different information-structural types interact with different information-theoretic factors in similar ways. For instance, could it be that the prosodic cues for new-information vs. corrective focus differ in terms of whether they are sensitive to word frequency vs. contextual probability?

Since prior work has found that the effect of givenness on duration is stronger when the repeated words are high-frequency (Baker & Bradlow, 2009), we hypothesized that the prosodic effect of information structure would be stronger in words with low informativity in the information-theoretic dimensions. In other words, the prosodic cues for information structure might be weakened when other factors – such as information-theoretical properties – also demand prosodic prominence. Building on Baker and Bradlow (2009), our study explored effects of word frequency and narrow new-information focus. We also looked at the effects of another information-theoretic factor, namely contextual probability, as well as another type of information structure, namely narrow corrective focus. Including multiple factors of each kind of informativity allowed us to investigate the potentially complex interactions among them. Specifically, we expected that narrow focus would be prosodically distinct from wide focus when the target word is highly frequent and/or highly contextually probable (i.e. has low information-theoretic informativity). In contrast, when the target word is low-frequency and/or low-probability (i.e. has high information-theoretic informativity), we predicted that the prosodic distinctions between narrow and wide focus might be weakened or even absent: prosodic reflexes of information structure might be observed in only one or perhaps in neither of the two narrow-focus conditions. If these predictions are borne out, we can then look into whether different information-structural types (i.e. corrective vs. new) could react differently to different information-theoretic factors (i.e. lexical frequency vs. contextual probability).

In addition to the general trends among speakers, the discussion in Section 1.4 shows that speakers differ in their acoustic realization of prosody. As there is not a lot of prior work focusing on individual differences in sentence prosody, we first wanted to see, on a general level, whether our results fit with the previous findings that sentence prosody is susceptible to speaker-specific effects. We then also looked more closely at whether and how individual differences manifested themselves in the prosodic encoding of informativity. Roughly speaking, we expected individual differences in all aspects investigated, because existing research on other prosody-related topics (as discussed in the preceding sections) has found both qualitative and quantitative variability among the participants of a study, in terms of the range and characteristics of cues a participant produces along an acoustic dimension as well as the size and direction of acoustic differences that a participant produces to signal a linguistic contrast (Andreeva, Barry & Steiner 2007; Dahan & Bernard, 1996; Fougeron & Kewley-Port, 2007; Krahmer and Swerts 2001; Niebuhr, D'Imperio, Fivela & Cangemi, 2011; Trouvain & Grice, 1999). Specifically, we expected our participants to differ in whether they made distinctions between narrow and wide focus in a given information-theoretic condition, whether they

increased or decreased prosodic prominence for a given region of the sentence, to what extent they vary prosodic prominence to convey the informativity of a word, as well as the overall prosody of their utterances.

In terms of the acoustic correlates of prosodic prominence, we focused on (i) the *shape of an f0 contour* and (ii) the *size of excursions in an f0 contour* (which will be called 'f0 range' henceforth). We chose f0 because it is an acoustic dimension that has been extensively studied in the information-structural tradition yet not much so in the information-theoretic tradition. In other words, by conducting this study, we also hoped to provide further evidence for the effects of information-theoretic factors on f0. Furthermore, because there are studies showing that intonational categories (e.g. H*, L+H*) do not necessarily map straightforwardly onto focus types (e.g. Katz & Selkirk, 2011; Krahmer & Swerts, 2001, Watson, Tanenhaus, & Gunlogson, 2008), we did not take an intonational-phonological approach (e.g. Ladd, 1996; Pierrehumbert and Hirschberg, 1990). Based on previous research, a good indicator of narrow focus seems to be a relatively exhaustive use of the acoustic space. In the f0 dimension, as mentioned in Section 1.1, it has been found that narrow focus differs from wide focus in having greater f0 protrusion or higher f0 on the narrowly focused element, greater f0 compression or sharper f0 fall following the focused element, and lower f0 preceding the focused element (Breen et al., 2010; Brown, 1983; Cooper et al, 1985; Couper-Kuhlen, 1984; Eady & Cooper, 1986; Katz & Selkirk, 2011). Therefore, we quantitatively measured both f0 shapes and f0 ranges, which presumably would capture the level of prominence in the f0 dimension.

In our study, the object of a sentence is the narrowly focused word in the discourse. Therefore, we expected narrow focus to influence prosody in the sentence region containing the object and the words immediately preceding and following it. Specifically, we predicted that the f0 movement of this sentence region would be bigger, or at least not smaller, in the narrow-focus conditions than the wide-focus condition. Also, we expected that individual participants would differ in the f0 shapes and ranges they produced in general, and the sizes and directions of differences they produced for information-structural distinctions.

2. Experiment

We conducted a production study with an interactive set-up. An earlier, abbreviated discussion of this experiment is available in Ouyang & Kaiser (2014), where we discuss some of the group results but do not explore any issues related to individual variation. Each trial consisted of a read-aloud task and a subsequent selection task. In both tasks, participants interacted with a partner, who was a lab assistant. The read-aloud task provided the critical recordings: the target sentences were produced by the participants during the read-aloud task. The selection task was included to engage both people in the read-aloud task: paying attention to what the other person said in the read-aloud task was necessary to successfully perform the selection task. (We do not

discuss the selection task in detail here because it is not relevant for the results, but people essentially had to pick the correct items from a larger array).

2.1. Design and procedures

Participants worked with a partner in reading aloud sentence pairs. Each sentence pair consisted of a question spoken by the partner (Sentence A) and a response (the critical sentence) spoken by the participant (Sentence B), as shown in (1-3) below. Participants saw Sentence B on a computer screen when it was their turn to speak. The target sentences (Sentence B on target trials) are transitive clauses with the following structure: a third-person plural pronoun subject, a simple past tense verb, an object, and a prepositional phrase indicating a location. The critical word we focus on is the object of each target sentence (e.g. *balls*). The experiment had 48 target trials; each participant encountered four items in each condition and did not see any item more than once. A full list of the target sentences can be found in Appendix 1. There were also 48 filler trials. The dependent variable we measured was the f0 values of an utterance.

(1) NARROW CORRECTIVE FOCUS
A: I heard that Dawn and Alice got gloves at the sports store.
B: No, they got [balls]$_{\text{CORRECTIVE FOCUS}}$ at the sports store.

(2) NARROW NEW-INFORMATION FOCUS
A: What did Rachel and Carolyn get at the sports store?
B: They got [balls]$_{\text{NEW-INFO FOCUS}}$ at the sports store.

(3) WIDE/VP FOCUS
A: What did Angela and Joyce do?
B: They [got balls at the sports store]$_{\text{NEW-INFO FOCUS}}$.

To investigate whether information-theoretic factors interact with information structure in shaping the prosody of an utterance, we manipulated (i) the lexical frequency of the object noun, (ii) whether the object was probable in the context of the preceding verb and the following location, and (iii) the object's informational-structural status in relation to the question. Thus, a within-subject design with three independent variables was implemented: **(i)** word frequency (with two levels: high or low frequency), **(ii)** contextual probability (with two levels: high or low probability), and **(iii)** focus type (with three levels: narrow corrective focus, narrow new-information focus, or wide/VP focus).

We manipulated the *focus type* of the critical noun by means of the question asked by the partner, as shown in (1)-(3). In the wide/VP focus condition (ex. 3), the question asks about the content of the entire verb phase (i.e., *what did X do?*), and the answer spoken by the participant provides this information. Thus, the whole VP (e.g., *got balls at the sports store*) is new information. In the narrow new-information focus condition (ex. 2), in contrast, the question asks for the object

of the transitive verb, and therefore only the object is new information in this condition. Finally, in the narrow corrective focus condition (ex. 1), the partner makes a statement where the object is incorrect (as signaled to the participant by the sentence on their screen), and thus the object in the participant's response is correctively focused.

The *contextual probability* of the critical words was estimated through a web-based norming study. Four verb-location contexts and eight objects were ultimately selected for the target sentences, as shown in Table 2. Each of the eight target nouns functioned as a probable object in some contexts and an improbable object in other contexts. This allows us to ensure that any effects of contextual probability cannot be attributed to idiosyncratic properties of specific nouns. Another four nouns were selected to be the 'incorrect' objects in the question that elicited corrective focus (e.g., *gloves* in ex. 1). These nouns had a contextual probability between the high-probability and low-probability critical words and a word frequency between the high-frequency and low-frequency critical words.

The *word frequency* of the object nouns was determined according to the SUBTLEXus database (Brysbaert & New, 2009). SUBTLEXus provides word frequency measures on the basis of American English subtitles, and contains 51 million words in total. The critical words in the high-frequency conditions ranged in a frequency from 67.76 to 40.16 per million, while those in the low-frequency conditions ranged in a frequency from 13.22 to 0.41 per million, as shown in Appendix 2.

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

Table 2: Manipulation of word frequency and contextual probability

2.2. Participants

Sixteen native speakers of American English participated in this study. All participants, 11 female and 5 male, were students at the University of Southern California. Two lab assistants interacted with participants in this study. Both lab assistants were female native speakers of American English and students at the University of Southern California.

2.3. Data analysis

768 utterances were collected from the 16 participants, each producing 48 target responses. Out of the full set of data, 43 utterances (5.6%) were not included in the data analysis, due to speech errors (16 utterances), disfluencies (6 utterances) and technical issues with the audio recording (21 utterances).

F0 measurements were obtained using the STRAIGHT algorithm (Kawahara, de Cheveigne & Patterson, 1998) through the VoiceSauce program (Shue, Keating, Vicenik & Yu, 2011). The raw f0 values were then smoothed (*smoothn* in MATLAB: Garcia, 2010) to remove f0 tracking errors and segmental effects. The smoothed values were then converted into a semitone scale, as semitones reflect pitch perception better than the Hertz scale (e.g. Nolan, 2003). Finally, the data were normalized by subject using z-scores, to factor out individual differences in f0 registers (e.g., women usually have wider and higher registers of f0 than men). The z-scores represented each data point in terms of its number of standard deviations above or below the mean across all utterances produced by a given speaker.

To investigate whether different levels of word frequency and contextual probability influence the prosodic encoding of narrow focus in different ways, we examined the effects of narrow focus in the four conditions of word frequency and contextual probability separately: high-frequency words in high-probability contexts, high-frequency words in low-probability contexts, low-frequency words in high-probability contexts, and low-frequency words in low-probability contexts[2]. As the prosodic effects of narrow focus were expected in the focused word and the words immediately before and after it (see Section 1.5 for the predictions of this study), we examined these regions of a sentence. Specifically, we analyzed f0 shapes and ranges for the following three intervals: Pre-Focus (verb), Focus (object), and Post-Focus (the region from preposition to article), from Pre-Focus to Focus (verb and object), and from Focus to Post-Focus (the region from object to article)[3]. F0 ranges were calculated by subtracting the minimum f0 from the maximum f0 in each interval. Since the f0 measurements have been normalized based on a given speaker's f0 register, a larger f0 range indicates that the participant was employing a bigger proportion of his or her f0 register.

To examine the *shapes of f0 contours*, we used a smoothing spline ANOVA approach. The smoothing spline ANOVA fits regression to continuous data to test differences between curves (Gu, 2002). We plotted the best-fitted curves with 95% confidence intervals (1.96 standard errors). The best-fitted values in a regression analysis can be interpreted as representing the

[2] We did not directly compare different levels of word frequency or contextual probability, because identical sentences existed only between different types of focus. This is an intrinsic property of the design, due to the manipulation of word frequency and contextual probability.
[3] We did not statistically analyze the head noun of the prepositional phrase because it was at the end of a sentence, where f0 varied considerably due to factors outside the scope of this study such as dialects (e.g. Ching, 1982) and turn transition cues (e.g. Wennerstrom and Siegel, 2003).

average patterns of the data being modeled. Two conditions can be considered as being significantly different if the 95% confidence intervals of their best-fitted values do not overlap. Similar approaches have been used for other kinds of continuous data in phonetics, such as tongue shapes (e.g. Davidson, 2006) and formants (e.g. Baker, 2006). We first extracted 20 data points with equal time spacing from each of the three consecutive intervals: Pre-Focus, Focus and Post-Focus. Mixed-effects smoothing spline ANOVA models were then performed with Focus Type, Time and their Interaction as fixed effects (*gss* in R: Gu, 2013). In the analysis of group patterns (presented in Section 3), Subject and Item were included as random intercepts. For the analysis of individual patterns (presented in Section 4.1), models were performed on each participant's data separately, and Item was included as a random intercept.

For *f0 ranges*, mixed-effects models were conducted on f0 ranges (*lme4* in R: Bates, Maechler, Bolker and Walker, 2014; *lmerTest* in R: Kuznetsova, Brockhoff, and Christensen, 2015). In the analysis of group patterns (presented in Section 3), Focus Type was included as a fixed effect, and Subject and Item were included as random effects. When specifying the structure of random effects, we started with a full model (i.e. including intercepts and slopes for Subject and Item), and if it failed to converge, we reduced the Subject slopes and/or the Item slopes until the model converged. All the group models that converged had random intercepts for Subject and Item. In the analysis of individual patterns, Subject was included as a fixed effect and Item was included as a random effect when we looked at a speaker's overall f0 ranges regardless of the condition (presented in Section 4.1). This model had both a random intercept and random slopes for Item. Finally, for the directionality and magnitude of differences in f0 ranges between conditions (presented in Section 4.2), we mainly focused on descriptive statistics, because the numbers of observations became relatively low when the data were split into small subsets by both subject and condition.

3. Group results

Overall, the predictions outlined in Section 1.5 about the general trends were borne out, as can be seen in Figures 1-4, which shows the smoothing spline ANOVA results. In terms of **f0 shapes**, the three types of focus do not significantly differ in the Pre-Focus interval (the first section marked on the x-axis). Significant differences in f0 shapes start emerging towards the end of the Focus interval (the middle section marked on the x-axis) and continue for most of the Post-Focus interval (the last section marked on the x-axis). Narrow corrective focus (solid lines) and narrow new-information focus (dashed lines) have a steeper f0 drop than wide focus (dotted lines) in some cases, depending on the narrowly-focused word's frequency and contextual probability. More specifically, when the word is high-frequency and occurs in a probable context (*got <u>balls</u> at the sports store*), both types of narrow focus differ significantly from wide focus (Figure 1, labelled 'High Freq + High Prob'). However, when a high-frequency word is focused in an improbable context (*got <u>fish</u> at the sports store*), only new-information focus differs significantly

from wide focus; corrective focus patterns with wide focus (Figure 2, 'High Freq + Low Prob'). In contrast, when the word is lexically infrequent but contextually probable (*got cleats at the sports store*), corrective focus differs significantly from wide focus; new-information focus does not (Figure 3, 'Low Freq + High Prob'). Finally, neither type of narrow focus differs from wide focus when it is an infrequent word focused in an improbable context (*got toys at the sports store*, Figure 4, 'Low Freq + Low Prob').

The analysis of **f0 ranges** finds parallel patterns to the above results of f0 shapes. There are no significant differences in f0 ranges when the Pre-Focus and Focus intervals are analyzed either jointly (i.e. treated as one region) or separately (t's < 1.723, p's > 0.086). The interaction between word frequency, contextual probability and focus types appears when the Post-Focus interval is analyzed alone or jointly with the Focus interval. In the condition of *lexically frequent and contextually probable words,* both types of narrow focus have significantly larger f0 ranges than wide focus (t's > 2.524, p's < 0.05, except for new-information focus in the Post-Focus interval: t = 1.458; p = 0.147). In the condition of *lexically frequent but contextually improbable words*, only new-information focus has larger f0 ranges than wide focus (t's > 1.994, p's < 0.05); corrective focus does not (t's < 1.650, p's > 0.100). In contrast, for *low-frequency but high-probability words*, corrective focus has larger f0 ranges than wide focus (t's > 2.159, p's < 0.05); new-information focus patterns with wide focus (t's < 1.091, p's > 0.276). Lastly, neither type of narrow focus differs from wide focus when *low-frequency words are focused in low-probability contexts* (t's < 1.366, p's > 0.173).[4]

Figure 1: High Freq + High Prob

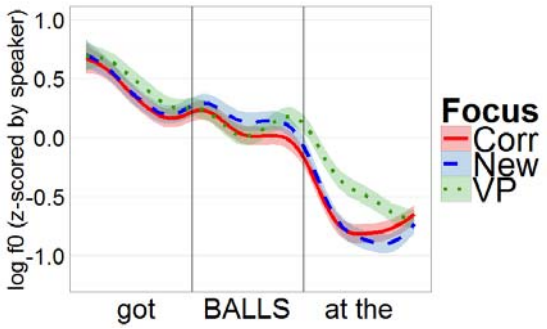


Figure 3: Low Freq + High Prob

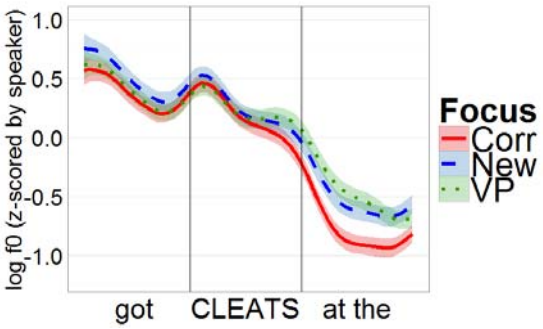
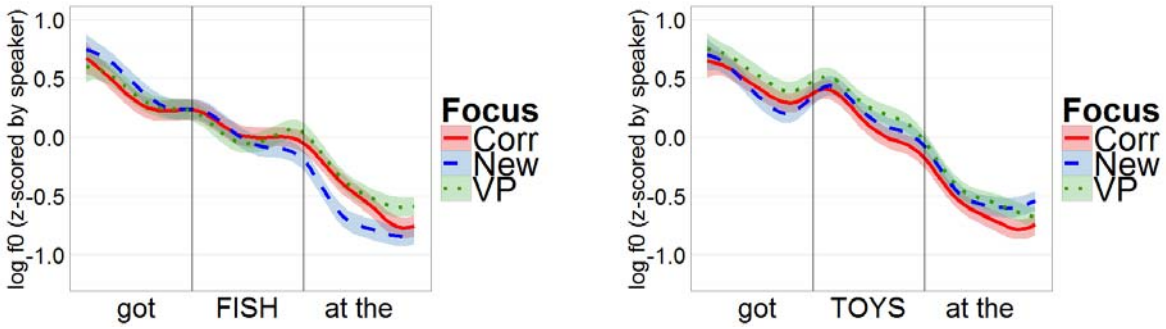


Figure 2: High Freq + Low Prob

Figure 4: Low Freq + Low Prob

[4] Here we do not report statistics for the Focus and Post-Focus intervals separately, due to reasons of readability, and more importantly because we do not think that this distinction (i.e. whether it is the Focus or Post-Focus interval which shows significant differences) is theoretically relevant for the claims we are making in this paper.



Figures 1-4: Best-fitted curves with 95% confidence intervals for the f0 values (in semitone, standardized by speaker) in the pre-focus, focus and post-focus regions of an utterance.

As a whole, we find that narrow focus brings greater prosodic prominence than wide focus, but this effect disappears under certain conditions of word frequency and contextual probability. Specifically, narrow corrective focus differs from wide focus only when the word carrying corrective information in narrow focus is probable in its sentence context. Conversely, new-information focus differs from wide focus when the word carrying new information in narrow focus is a frequent word. This suggests that the prosodic prominence associated with information structure is modulated by word frequency and contextual probability.

4. Individual results

In the previous section, we summarized the overall patterns when all participants are investigated as a group. Let us now explore whether and how individual participants differ from one another. In this section, we will first look at the overall prosody of individual speakers, focusing on f0 shapes and ranges (Section 4.1). Then, we will examine the different experimental conditions, to see how individual speakers produce different types of focus in different conditions of word frequency and contextual probability (Section 4.2).

4.1. Overall prosody of utterances

Overall, in terms of general prosodic patterns, speaker-specific variation occurs both qualitatively and quantitatively. Between-subject variability and within-subject consistency were observed in both the shapes of f0 contours and the ranges of f0 values.

First, *the shapes of f0 contours* vary greatly from participant to participant. In a given condition, participants differ in the number, locations and relative height of the f0 peaks and valleys that they produce in an utterance. To illustrate the extent of variability, we plotted a sample of five participants whose f0 shapes are clearly distinct from one another. Figure 5 shows the observed f0 contours produced by these participants for new-information, frequent words that are

narrowly-focused in probable contexts (e.g., *What did Rachel and Carolyn get at the sports store? They got* <u>*balls*</u> *at the sports store.*) We can see that participants 04 (triangles), 06 (dots) and 09 (dashes) all tend to produce a high tone on the focused word (i.e. *balls*) – thus showing overall consistency in this regard. However, their choices regarding the adjacent tones differ. Participant 06's utterances on average have a low tone preceding the high tone, participant 04's in general have another high tone preceding the high tone, and participant 09's seem to have a low tone following the high tone. Furthermore, participant 01 (squares) and participant 04 both show a clear tendency of declination, but participant 04's utterances have two high tones whereas participant 01's do not have apparent tone targets. Lastly, participant 07 (solid line) distinctively produces the focused word with a low tone. Such diversity is found among other participants and in other conditions as well.

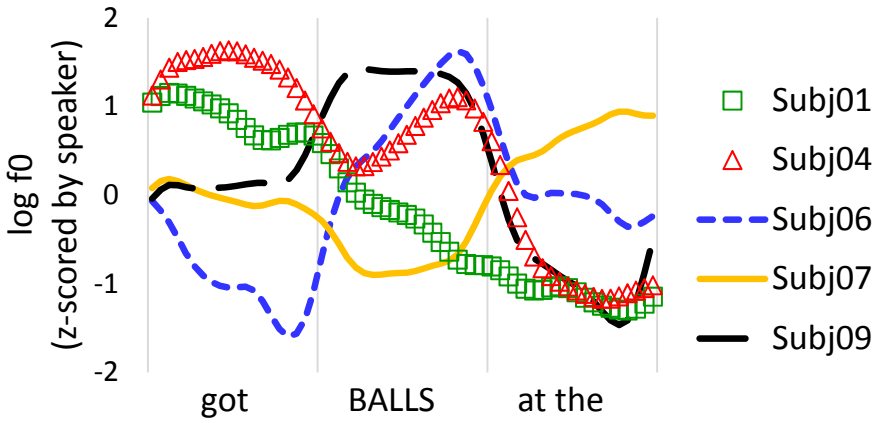


Figure 5: Observed mean f0 (in semitone, standardized by speaker) for participants 01, 04, 06, 07 and 09 in the narrow new-information focus, high word frequency and high contextual probability condition.

Although different participants produce different shapes of f0 contours, they show consistent patterns within their own utterances. For example, Figure 6 provides a glance at the observed f0 contours produced by participant 04 in all twelve experiment conditions. We can see that participant 04's utterances are quite similar to one another, regardless of the condition. To further illustrate the intra-subject consistency with better graphical legibility, Figure 7 shows the smoothing spline ANOVA results of three individual participants, including participant 04, in all four information-theoretic conditions. These three participants were chosen because they had strong preferences regarding f0 shapes. We can see that participant 01's utterances (top row) mostly follow a declination slope, although a low tone occasionally occurs around the end of the Focus interval. Participant 04 (middle row) consistently produces a high tone in the Pre-Focus interval and another high tone, downstepped, in the Focus interval, except there is sometimes a low tone proceeding and/or following the second high tone. Participant 06 (bottom row) generally produces a low tone in the Pre-Focus interval and a high tone in the Focus interval,

which is often followed by another low tone. Speaker-specific preferences of this sort are also found for most of the other participants in our data.

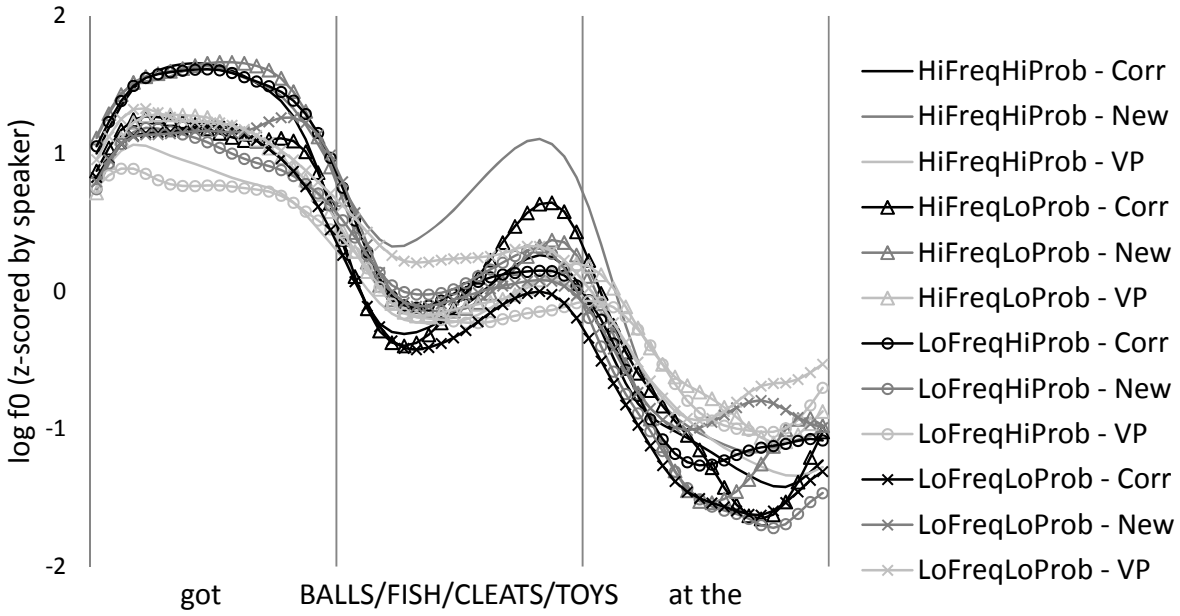


Figure 6: Observed mean f0 (in semitone, standardized by speaker) of participant 04 in all the experiment conditions.

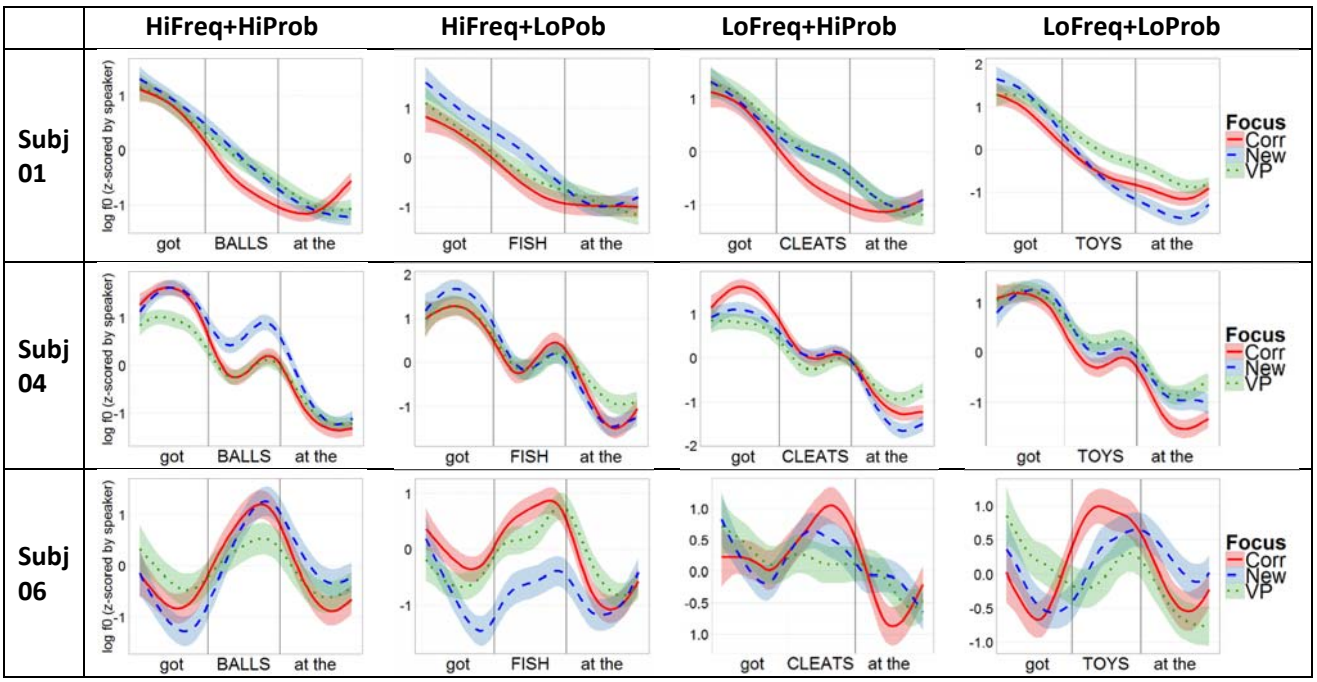


Figure 7: Best-fitted curves with 95% confidence intervals for the f0 values (in semitones, standardized by speaker) produced by participants 01, 04 and 06.

We also find speaker-specific effects in *the ranges of f0 values*. Some participants regularly employ a large proportion of their f0 register, while others regularly employ a small proportion of their f0 register. To illustrate, let us take a close look at the sentence region from the Pre-Focus interval to the Post-Focus interval. Figure 8 shows the average f0 ranges with 95% confidence intervals (1.96 standard errors) produced by individual participants. We can see that every participant differs from some other participant(s). Pairwise comparisons with the Bonferroni adjustment show that, between the sixteen participants, everyone significantly differs from at least two other people and as many as thirteen other people (p's<0.05). For example, participant 05, whose f0 ranges are largest on average (mean = 2.787) and the least variable among all participants (standard deviation = 0.512), differs from participants 01, 03, 04, 06, 07, and 09-16. On the other hand, participant 12, whose f0 ranges are smallest on average (mean = 1.582), differs from participants 02, 04, 05, 06, 08, 09, 11 and 16. Even participant 07, whose f0 ranges are the most variable among all participants (standard deviation = 1.253), differs from participants 02, 05, 08 and 11 (by being smaller). More details about other participants can be observed in Figure 8.

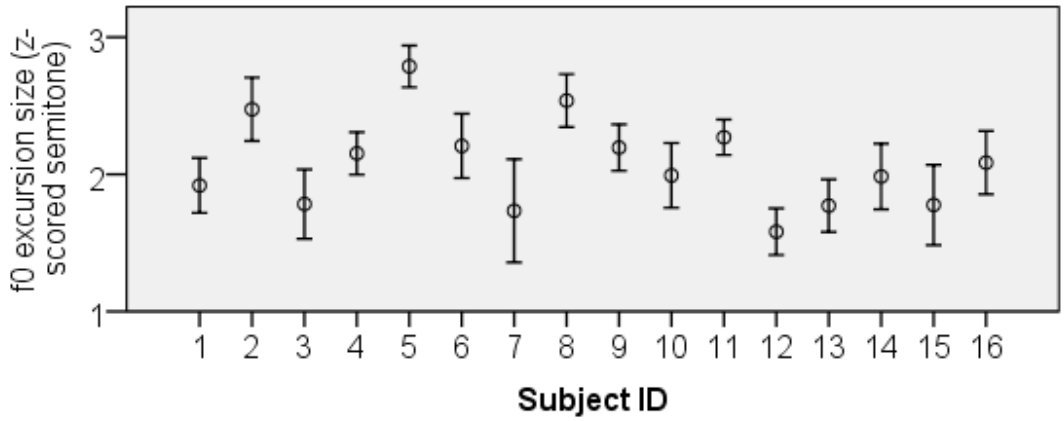


Figure 8: The observed f0 ranges (calculated from semitones standardized by speaker) with 95% confidence intervals for individual participants in the sentence region from the pre-focus interval to the post-focus interval. A larger f0 range indicates that the speaker employs a bigger proportion of his/her f0 register for this sentence region.

In sum, individual participants appear to be fairly different from one another, yet consistent within one's own utterances, in terms of the f0 shapes they adopt and how large a proportion of their f0 register they use. This suggests evidence for speaker-specific behavior in the overall prosodic patterns of utterances and the extent to which people utilize their vocal capacity to produce prosodic cues.

4.2. Prosodic encoding of informativity

Now that we have seen speaker-specific effects on the overall shapes and ranges of f0, let us move on to the individual differences in how their prosody reflects the informativity of linguistic elements. Since a given participant's f0 shapes are similar across the conditions, i.e., different types of focus and different levels of word frequency and contextual probability (see Section 4.1), only f0 ranges are of the interest in this subsection. To draw a direct comparison between the group trends and the individual patterns, we present the results of the sentence region from the Focus interval to the Post-Focus interval, where the group analysis finds significant differences (see Section 3).

First, we observe some between-subject variability in terms of the *direction* of distinctions between different kinds of information. As presented in Section 3, there are three main patterns when all sixteen participants are analyzed as a group: (i) wide focus has smaller f0 ranges than both types of narrow focus in the condition of *frequent and probable words*, (ii) narrow new-information focus has larger f0 ranges than narrow corrective focus and wide focus in the condition of *frequent but improbable words*, and (iii) narrow corrective focus has larger f0 ranges than narrow new-information focus and wide focus in the condition of *infrequent but probable words*. The analysis of individual participants finds each pattern in eight or nine people out of sixteen: pattern (i) is exhibited by participants 01, 02, 04, 05, 07, 10, 12, 14 and 15; pattern (ii) is exhibited by participants 01, 02, 06, 07, 12, and 14-16; pattern (iii) is exhibited by participants 04 and 07-13. In other words, only about half of the participants conform to the group trends regarding how information-structural types are differentiated in a given information-theoretic condition, and it is not the same individuals in every condition. However, it is worth noting that there are no alternative 'competitor' patterns – instead, the participants who do not match the overall group trends show a mix of patterns in the different conditions. Thus, although the overall group trends (as summarized in (i)-(iii) above) are not exhibited by everyone, they nevertheless constitute the clearest patterns that emerge from the data.

Participants also differ in the *magnitude* of the information-structural distinctions they make. To illustrate, Figure 9 shows the f0 ranges of individual participants in the condition of high word frequency and high contextual probability. It appears that some people make clearer distinctions than others. For example, the differences between wide and narrow focus are bigger in participants 07 and 15 than participants 04 and 14. Participants 07 and 15 use substantially larger f0 ranges for the utterances containing narrow focus than the utterances containing wide focus, whereas participants 04 and 14 differentiate these two kinds of utterances to a lesser degree. Similarly variable patterns are found in other conditions as well.

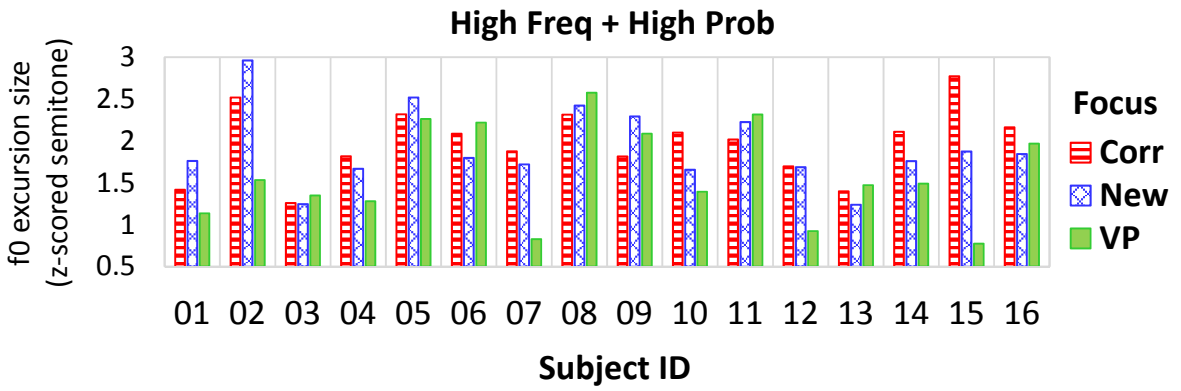


Figure 9: The observed f0 ranges (calculated from semitones standardized by speaker) in the sentence region from the Focus interval to the Post-Focus interval for individual participants in the condition of high word frequency and high contextual probability. A larger f0 range indicates that a bigger proportion of the speaker's f0 register is employed.

Let us now consider how internally-consistent speakers are in terms of the (i) directionality and (ii) magnitude of the information-structural distinctions that they produce. We find considerable trial-by-trial variation in the direction of the information-structural distinctions produced by individual participants (although the patterns reach significance in the group analysis). Particularly, there is little indication of interactions between speaker (i.e. who is speaking) and any of the informativity factors in terms of the *direction* of distinctions between different kinds of information. In other words, the overall group results also hold on the level of individual speakers, and it is generally not the case that, depending who the speaker is, one particular type of information would consistently lead to smaller (or larger) f0 ranges than another particular type of information.

Interestingly, if we look at the magnitude of these distinctions, we find more speaker-internal consistency. Some participants regularly produce much larger f0 ranges for one type of focus than another, while some others regularly produce only slightly larger f0 ranges for one type of focus than another. For example, let us take a close look at the participants who conform to more than one group trend: participants 01, 02, 04, 07, 10, 12, 14 and 15. It appears that they can be divided into two subgroups such that, across information-theoretic conditions, one subgroup consistently produces stronger cues for information-structural distinctions than the other subgroup. To illustrate, Figure 10 shows the differences in f0 ranges produced by the eight participants in the information-theoretic conditions where they conform to the group trends regarding the information-structural distinctions. These differences were calculated with respect to the group trend in each condition, i.e. patterns (i-iii) that we summarized towards the beginning of this subsection. Specifically, the bars for the high-frequency high-probability condition represent the differences between wide focus and the other two types focus (i.e. the f0

range in wide focus subtracted from the f0 ranges in new-information narrow focus and corrective narrow focus) based on pattern (i)), the bars for the high-frequency low-probability condition represent the differences between narrow new-information focus and the other two types of focus (i.e. the f0 ranges in wide focus and corrective focus subtracted from the f0 range in new-information narrow focus, based on pattern (ii)), and the bars for the low-frequency high-probability condition represent the differences between narrow corrective focus and the other two types of focus (i.e. the f0 ranges in wide focus and new-information focus subtracted from the f0 range in corrective narrow focus, based on pattern (iii)). For the participants who do not conform to all of these three group patterns (i.e. participants 1, 2, 4, 10, 14, and 15), we only calculated the differences in f0 ranges for the information-theoretic conditions where they do. Thus, we can see that participants 02, 07, 12 and 15 produce pattern (i) with larger differences than participants 01, 04, 10 and 14, participants 02, 07, 12 and 15 produce pattern (ii) with larger differences than participants 01 and 14, and participants 07 and 12 produce pattern (iii) with larger differences than participants 04 and 10. Essentially, in a given information-theoretic condition, participants 02, 07, 12 and 15 consistently use larger differences in f0 ranges than participants 01, 04, 10 and 14 for a given direction of information-structural distinctions. In general, this observation leads us to speculate that what matters (in terms of encoding and perceiving informativity) are not the absolute but rather the relative values.

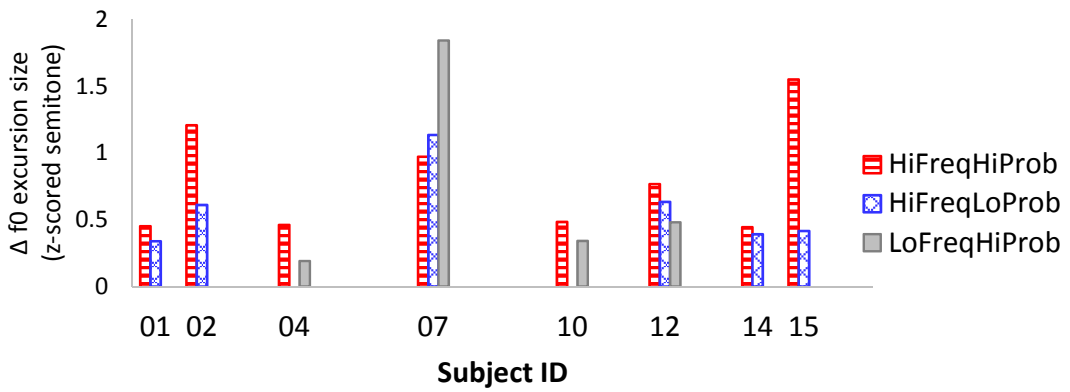


Figure 10: The observed differences in f0 ranges (calculated from semitones standardized by speaker) in the sentence region from the Focus interval to the Post-Focus interval for individual participants who conform to more than one group trend. The differences were calculated based on the group trend in each condition.

To sum up, when we look at individual differences in how speakers encode informativity prosodically, we find that about half of the speakers clearly exhibit the f0 range patterns that we observed for the group as a whole in terms of which conditions have larger vs. smaller f0 ranges, and the remaining speakers show more variable data. In terms of the magnitude of their f0 ranges, speakers are largely internally consistent, and our data suggests that speakers differ in how much they modulate f0 to signal informativity. Broadly speaking, this suggests that what matters in

terms of encoding information-theoretic notions prosodically are relative, not absolute, values – an observation which is in line with prior work on prosody and information structure.

5. General discussion

Our experiment investigates how information structure and information-theoretic properties interact in shaping the prosody of an utterance and how individual speakers differ in the overall prosody of utterances and the prosodic encoding of informativity. Existing studies have examined prosody from information-theoretic and information-structural perspectives, but the interaction between these two kinds of informativity factors has not been thoroughly investigated. In addition, prior work mostly focuses on the general trends among speakers, and little has been said about the differences between or within speakers. A better understanding of these issues is important because they are involved in fundamental questions regarding the functions and constraints of the prosodic system. In this section, we discuss how our results relate to these issues, and what their broader implications are.

Our results in Section 3 show that, when the **participants are analyzed as a whole,** the prosodic effects of information structure are modulated by information-theoretic factors. In particular, we find differential effects of contextual probability and word frequency on corrective narrow focus vs. new-information narrow focus. *Corrective narrow focus* results in significant f0 movement only when the word carrying corrective information is *probable* in the context. However, *new-information narrow focus* results in significant f0 movement only when the word carrying new information is a *frequent* word. When the narrowly focused word is lexically frequent and contextually probable, both types of narrow focus have greater f0 movement than wide focus. In contrast, when the narrowly focused word is infrequent and improbable, neither type of narrow focus type is distinguishable from wide focus. This fits with our prediction that the prosodic prominence associated with information structure would be weakened when other factors also demand prosodic prominence.

Taken together, these findings pose a challenge to the widespread view that narrow focus is (consistently) associated with greater prosodic prominence than wide focus. In fact, prior work on the phonetic realization of information structure suggests a prominence hierarchy, such that contrastive/corrective information is prosodically more marked than 'plain' new information, and new information in narrow focus is prosodically more marked than new information in wide focus (e.g., English: Breen et al. 2010; Katz & Selkirk 2011; German: Baumann et al. 2006; Mandarin Chinese: Ouyang & Kaiser, 2015; Xu 1999). To the contrary, we did not see this hierarchy in our data – we found that contextual probability and word frequency need to be considered in order to understand the relative prosodic prominence of different focus types. Interestingly, it seems that many existing studies have focused on relatively probable contexts and have not manipulated word frequency, which may explain the hierarchical relation

previously found between corrective focus, new-information focus and wide focus (i.e. narrow corrective > narrow new > wide new).

Consider a hypothetical study that has a mix of high-frequency and low-frequency words focused in probable contexts. Based on our results, in such a study: (a) corrective focus will have greater prominence than wide focus, since the contexts are probable, and (b) new-information focus will be less prominent than corrective focus and more prominent than wide focus, because frequent words pattern with the former but infrequent words pattern with the latter. These predictions are confirmed by a follow-up analysis where we pooled the conditions of word frequency and excluded the condition of high contextual probability. Using the approaches described in Section 2.3, we found significant differences in the Focus and Post-Focus intervals. The f0 movement was largest for corrective focus, second largest for new-information focus, and smallest for wide/VP focus. In other words, the common generalization about the prominence hierarchy between the three types of focus might be an epiphenomenon stemming from not controlling word frequency and using relatively probable contexts.

Here we will not further discuss why a word's information-theoretic properties interact with its information-structural status in the particular way we observed, since it is not the focus of this paper. Nevertheless, our findings highlight the importance of disentangling information structure and information-theoretic factors. To fully understand how prosody encodes informativity, it is necessary to integrate the work in the information-theoretic approach and the work in the information-structural approach (see Wagner & Watson, 2010: 933, for relevant discussion).

Let us now consider the **nature and extent of individual variation.** In this section, we will consider the shapes of f0 contours, the ranges of f0 values, the directionality of differences in f0 ranges (i.e. which conditions have larger/smaller f0 ranges than other conditions), and the magnitude of differences in f0 ranges (i.e. how much larger/smaller the f0 ranges are in one condition than another). As we saw in Section 4, if we look at the overall prosody and f0 ranges that speakers produce, abstracting away from informativity notions, we find that speakers differ from one another but are internally quite consistent. In other words, individual speakers have preferences with regard to the shapes of f0 contours and the ranges of f0 values, generally speaking. Then, when we look at how individual speakers encode informativity notions prosodically, we find that the group-level patterns regarding the directionality of information-structural distinctions are exhibited by many, but not all, speakers. Interestingly, when we look more closely at how internally consistent speakers are in this regard, we find that speakers show considerable internal variation in the directionality of distinctions they produce (i.e. whether a particular type of focus has larger or smaller f0 ranges than another particular type of focus). In contrast, in terms of the magnitude of distinctions they produce (i.e. how much larger or smaller the f0 ranges are in one particular type of focus than another), speakers are more internally consistent while, again, different from one another. Nevertheless, the group patterns are

statistically significant (in analyses that include subjects and items as random factors), and thus we conclude that they are still meaningful even in the face of individual variation.

As we noted in Section 3, the group analysis reveals three main patterns which highlight the interplay of information theory and information structure: (i) wide focus has smaller f0 ranges than both types of narrow focus in the condition of *frequent and probable words*, (ii) narrow new-information focus has larger f0 ranges than narrow corrective focus and wide focus in the condition of *frequent but improbable words*, and (iii) narrow corrective focus has larger f0 ranges than narrow new-information focus and wide focus in the condition of *infrequent but probable words*. We found that about half of the participants clearly exhibit these patterns. Importantly, there is no other 'competitor pattern' that emerges from the data, as the rest of the participants exhibit more than one other pattern (e.g. some make corrective focus the least prominent while others make new-information focus the least prominent, as can be seen in Figure 9).

Thus, we observe a set of patterns that a large subset of participants exhibits, and then other, seemingly highly variable, non-systematic patterns. It seems that speakers loosely follow principles determined by information-theoretic factors and information structure, and collectively show a systematic relationship between prosodic prominence and informativity. A related phenomenon has been found in the field of speech processing. Studies on accent prediction have argued that using speaker-dependent parameters does not substantially improve a model's performance in predicting whether a word receives an accent or not, because the variability in placing an accent or not between speakers is similar to that within a speaker (Badino & Clark,2007; Shriberg, Ladd, Terken & Stolcke, 1996; Yuan, Brenier & Jurafsky, 2005). Our results are consistent with these findings.

While the directions of differences in f0 ranges are closely tied to informativity factors, some other aspects of f0 – including the ranges of f0 values, the sizes of differences in f0 ranges, and the shapes of f0 contours – appear to show speaker-specific behavior. Given the multi-functionality of prosody, it is not surprising that these other f0 parameters do not supply strong cues for the particular factors we investigated. In terms of the range of f0 values in an utterance and the magnitude of fluctuations in f0 ranges across utterances, prior work has found that these aspects of f0 ranges can reflect the speaker's emotions and psychological traits. For example, sad, depressed, anxious, irritated, tense or fearful speech employs more limited f0 ranges than happy or angry speech (e.g. Johnstone & Scherer, 1999; Morley, van Santen, Klabbers & Kain, 2011). Furthermore, children and young adults with autistic spectrum disorders use more exaggerated f0 ranges than individuals with typical development (e.g. Hubbard & Trauner, 2007; Paul, Bianchi, Augustyn, Klin & Volkmar, 2008; Sharda et al., 2010). Thus, it is likely that the speaker-specific patterns regarding f0 ranges observed in this study correlate with individual participants' mood or personal characteristics.

Similarly, f0 shapes have been shown to convey many other kinds of pragmatic meanings that are not investigated in this study, such as the speaker's beliefs or the relationship between an utterance and a subsequent one (e.g. Pierrehumbert & Hirschberg, 1990; Ward & Hirschberg, 1985, 1986). Due to the nature of our experiment (i.e. reading aloud sentence pairs), the stimuli were underspecified in these aspects and open for the participant's own interpretations. Therefore, the presence of speaker-specific patterns in f0 shapes might imply that individual speakers have preferences regarding how to fill in unspecified details at the pragmatic level. This is an interesting question that would benefit from future work.

Thus, based on our results, it appears that f0 shapes are less informative than f0 ranges in distinguishing the three information-structural types of interest, namely corrective narrow focus, new-information narrow focus, and wide focus. F0 shapes differentiate these three types of focus when we look at all speakers as a whole, but not when we look at each speaker individually. In contrast, the directions of differences in f0 ranges distinguish focus types at *both* the group level and the individual level. This suggests that f0 ranges may have a greater contribution than f0 shapes to the prosodic marking of information structure. We leave this question open for future work.

In sum, this study contributes to our understanding of individual differences, providing empirical evidence for inter- and intra-speaker variability in the prosodic encoding of informativity. Our results are consistent with previous observations that prosody can exhibit speaker-specific behavior. Furthermore, we show that apparent differences among the participants in a study do not necessarily constitute stable speaker-specific patterns. Instead, the prosodic dimensions that do not show participants' individual preferences may be the key dimensions that reflect the linguistic distinctions in question (e.g. the direction of differences in f0 ranges in this study). In addition, we discuss possible explanations for speaker-specific behavior in the prosodic dimensions we investigate. Prosody appears to be highly multi-functional and tolerant of idiosyncrasies to a considerable extent.

6. Conclusions

On the basis of the psycholinguistic production study reported in this paper, we can draw three main conclusions. *First*, information structure and information-theoretic factors interact in influencing an utterance's prosody. Our results show that word frequency modulates the prosodic effect of new-information focus (see also Baker & Bradlow, 2009), whereas contextual probability modulates the prosodic effect of corrective focus. *Second*, our findings suggest the presence of speaker-specific behavior in prosody. Speakers have individual preferences regarding the prosodic patterns of utterances and the magnitude of prosodic cues for informativity. *Third*, we did not see signs of speaker-specific behavior in the directions of

prosodic distinctions between information categories – in other words, this seems to be a key dimension where English speakers show consistent behavior in terms of how informativity related factors are encoded in prosody. In sum, this work contributes to our understanding of prosody by providing empirical evidence for the interaction between word frequency and new-information focus, the interaction between contextual probability and corrective focus, as well as the nature and extent of speaker-specific variation. Our findings highlight the importance of disentangling information structure and information-theoretic factors and examining both inter- and intra-speaker variability.

Acknowledgements

Earlier version of this work was presented at the 4th *International Summer School 2013 on Speech Production and Perception: Speaker-Specific Behavior,* the 27th Annual CUNY Conference on Human Sentence Processing (2014, Columbus, Ohio, USA), the 38th Annual Penn Linguistics Conference in 2014, and the 36th Annual Meeting of the Cognitive Science Society (2014, Quebec City, Quebec, Canada). We thank the audience members for their valuable comments and suggestions. Thanks also go to the USC Language Processing Lab group for feedback during the development of this project. Last but not least, we thank the editors of this book and three anonymous reviewers, whose comments and suggestions greatly enhanced this chapter.

References

Allen, J. Sean, Joanne L. Miller, and David DeSteno. 2003. Individual talker differences in voice-onset-time. *Journal of the Acoustical Society of America* 113:544–552.

Andreeva, Bistra, William J. Barry, and Ingmar Steiner. 2007. Producing phrasal prominence in German. In *Proceedings of the 16th International Congress of Phonetic Sciences*, 1209-1212.

Aylett, Matthew, and Alice Turk. 2004. The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence and duration in spontaneous speech. *Language and Speech* 47:31–56.

Badino, Leonardo, and Robert A. J Clark. 2007. Issues of optionality in pitch accent placement. In *Proceedings of the 6th ISCA Workshop on Speech Synthesis*, 252-257.

Baker, Adam. 2006. Quantifying diphthongs: A statistical technique for distinguishing formant contours. Paper presented at New Ways of Analyzing Variation (NWAV) 35, Columbus, OH.

Baker, Rachel E., and Ann R. Bradlow. 2009. Variability in word duration as a function of probability, speech style, and prosody. *Language and Speech* 52:391–413.

Bates, Douglas, Martin Maechler, Ben Bolker, and Steven Walker. 2014. *lme4: Linear mixed-effects models using Eigen and S4*. R package version 1.1-7.

Baumann, Stefan, Martine Grice, and Susanne Steindamm. 2006. Prosodic marking of focus domains-categorical or gradient. In *Proceedings of Speech Prosody 2006*, 301-304.

Bell, Alan, Jason M. Brenier, Michelle Gregory, Cynthia Girand, and Dan Jurafsky. 2009. Predictability effects on durations of content and function words in conversational English. *Journal of Memory and Language* 60:92–111.

Bell, Alan, Dan Jurafsky, Eric Fosler-Lussier, Cynthia Girand, Michelle Gregory, and Daniel Gildea. 2003. Effects of disfluencies, predictability, and utterance position on word form variation in English conversation. *Journal of the Acoustical Society of America* 113:1001–1024.

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

Brown, Gillian. 1983. Prosodic structure and the given/new distinction. In Anne Cutler and D. Robert Ladd (Eds.), *Prosody: Models and Measurements*. Springer Science and Business Media.

Brysbaert, Marc, and Boris New. 2009. Moving beyond Kucera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods* 41:977–990.

Calhoun, Sasha. 2010. How does informativeness affect prosodic prominence? *Language and Cognitive Processes* 25:1099-1140.

Chen, Yiya, and Braun, Bettina. 2006. Prosodic realization of information structure categories in Standard Chinese. In *Proceedings of Speech Prosody 2006*.

Ching, Marvin. K. L. 1982. The question intonation in assertions. *American Speech*, 95-107.

Cho, Taehong and Patricia A. Keating. 2001. Articulatory and acoustic studies on domain-initial strengthening in Korean. *Journal of Phonetics* 29:155–190.

Clopper, Cynthia G., and Janet B. Pierrehumbert. 2008. Effects of semantic predictability and regional dialect on vowel space reduction. *Journal of the Acoustical Society of America* 124:1682-1688.

Cooper, William E., Stephen J. Eady, and Pamela R. Mueller. 1985. Acoustical aspects of contrastive stress in question-answer contexts. *Journal of Acoustical Society of America* 77:2142–2156.

Couper-Kuhlen, Elizabeth. 1984. A new look at contrastive intonation. In Richard J. Watts and Urs Weidman (Eds.), *Modes of Interpretation: Essays Presented to Ernst Leisi on the Occasion of His 65th Birthday*. Tübingen: Gunter Narr Verlag.

Dahan, Delphine & Jean-Marc Bernard. 1996. Interspeaker variability in emphatic accent production in French. *Language and Speech* 39:341-374.

Davidson, Lisa. 2006. Comparing tongue shapes from ultrasound imaging using smoothing spline analysis of variance. *Journal of the Acoustical Society of America* 120:407–415.

Dik, Simon C. 1997. *The Theory of Functional Grammar*. Berlin: Mouton De Gruyter.

Eady, Stephen J., and William E. Cooper. 1986. Speech intonation and focus location in matched statements and questions. *Journal of the Acoustical Society of America* 80:402-415.

Ferguson, Sarah Hargus. 2004. Talker differences in clear and conversational speech: Vowel intelligibility for normal hearing listeners. *Journal of the Acoustical Society of America* 116:2365-2373.

Ferguson, Sarah Hargus, and Diane Kewley-Port. 2007. Talker differences in clear and conversational speech: Acoustic characteristics of vowels. *Journal of Speech, Language, and Hearing Research* 50:1241-1255.

Fougeron, Cécile, and Patricia A. Keating. 1997. Articulatory strengthening at edges of prosodic domains. *The Journal of the Acoustical Society of America* 101:3728–3740.

Fowler, Carol. A., and Johnathan Housum. 1987. Talkers' signaling of "new" and "old" words in speech and listeners' perception and use of the distinction. *Journal of Memory and Language* 26:489-504.

Garcia, Damien. 2010. Robust smoothing of gridded data in one and higher dimensions with missing values. *Computational Statistics and Data Analysis* 54:1167–1178.

Gregory, Michelle L., William D. Raymond, Alan Bell, Eric Fosler-Lussier, and Dan Jurafsky. 1999. The effects of collocational strength and contextual predictability in lexical production. *Chicago Linguistic Society* 35:151-166.

Gu, Chong. 2002. *Smoothing Spline ANOVA Models*. New York: Springer.

Gu, Chong. 2014. *gss: General Smoothing Splines*. R package version 2.1-4.

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

Hay, Jennifer, Paul Warren, and Katie Drager. 2006. Factors influencing speech perception in the context of a merger-in-progress. *Journal of Phonetics* 34:458-484.

Hubbard, Kathleen, and Doris A. Trauner. 2007. Intonation and emotion in autistic spectrum disorders. *Journal of Psycholinguistic Research* 36:159-173.

Johnstone, Tom, and Klaus R. Scherer. 1999. The effects of emotions on voice quality. In *Proceedings of the 16th International Congress of Phonetic Sciences*, 2029-2032.

Katz, Jonah, and Elisabeth Selkirk. 2011. Contrastive focus vs. discourse-new: Evidence from phonetic prominence in English. *Language* 87:771–816.

Kawahara, Hideki, Alain de Cheveigne, and Roy D. Patterson. 1998. An instantaneous-frequency-based pitch extraction method for high-quality speech transformation: Revised TEMPO in the STRAIGHT-suite. In *Proceedings of the 5th International Conference on Spoken Language Processing*, 1367–1370.

Krahmer, Emiel, and Marc Swerts. 2001. On the alleged existence of contrastive accents, *Speech Communication* 34:391-405.

Krivokapić, Jelena, and Dani Byrd. 2012. Prosodic boundary strength: An articulatory and perceptual study. *Journal of Phonetics* 40:430–442.

Kuznetsova, Alexandra, Per Bruun Brockhoff, and Rune Haubo Bojesen Christensen. 2015. *lmerTest: Tests in Linear Mixed Effects Models*. R package version 2.0-25.

Ladd D. Robert. 1996. *Intonational Phonology*. Cambridge: Cambridge University Press.

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

Loakes, Deborah, and Kirsty McDougall. 2010. Individual variation in the frication of voiceless plosives in Australian English: A study of twins' speech. *Australian Journal of Linguistics* 30:155–181.

Morley, Eric, Jan van Santen, Esther Klabbers, and Alexander Kain. 2011. F0 range and peak alignment across speakers and emotions. In *Proceedings of 2011 IEEE International Conference on Acoustics, Speech and Signal Processing*, 4952-4955.

Munson, Benjamin, and Nancy Pearl Solomon. 2004. The influence of phonological neighborhood density on vowel articulation. *Journal of Speech, Language, and Hearing Research* 47:1048-1058.

Niebuhr, Oliver, Mariapaola D'Imperio, Barbara Gili Fivela, and Francesco Cangemi. 2011. Are there "shapers" and "aligners"? Individual differences in signalling pitch accent category. In *Proceedings of the 17th International Congress of Phonetic Sciences*, 120–123.

Nolan, Francis. 2003. Intonational equivalence: an experimental evaluation of pitch scales. In *Proceedings of the 15th International Congress of Phonetic Sciences*, 771–774.

Ouyang, Iris Chuoying, and Elsi Kaiser. 2015. Prosody and information structure in a tone language: An investigation of Mandarin Chinese. *Language, Cognition and Neuroscience* 30:57-72.

Ouyang, Iris Chuoying, and Elsi Kaiser. 2014. Prosodic encoding of informativity: Word frequency and contextual probability interact with information structure. In *Proceedings of the 36th Annual Meeting of the Cognitive Science Society*, 1120–1125.

Pan, Shimei, and Julia Hirschberg. 2000. Modeling local context for pitch accent prediction, In *Proceedings of the 38th Annual Conference of the Association for Computational Linguistics*, 233–240.

Paul, Rhea, Nancy Bianchi, Amy Augustyn, Ami Klin, and Fred R. Volkmar. 2008. Production of syllable stress in speakers with autism spectrum disorders. *Research in Autism Spectrum Disorders* 2:110-124.

Pierrehumbert, Janet B., & Julia Hirschberg. 1990. The meaning of intonational contours in the interpretation of discourse. In Philip R. Cohen, Jerry L. Morgan, and Martha E. Pollack (Eds.), Intentions in communication. Cambridge, Massachusetts: MIT Press.

Pitrelli, John F. 2004. ToBI prosodic analysis of a professional speaker of American English. In *Proceedings of Speech Prosody 2004*.

Pluymaekers, Mark, Mirjam Ernestus, and R. Harald Baayen. 2005a. Lexical frequency and acoustic reduction in spoken Dutch. *Journal of the Acoustical Society of America* 118:2561–2569.

Pluymaekers, Mark, Mirjam Ernestus, and R. Harald Baayen. 2005b. Articulatory planning is continuous and sensitive to informational redundancy. *Phonetica-Basel* 62(2-4):146.

Prince, Ellen. 1992. The ZPG letter: Subjects, definiteness, and information-status. In William C. Mann and Sandra A. Thompson (Eds.), *Discourse Description: Diverse Analyses of a Fund-Raising Text*. Philadelphia: John Benjamins.

Rietveld, Toni, Joop Kerkhoff, and Carlos Gussenhoven. 2004. Word prosodic structure and vowel duration in Dutch. *Journal of Phonetics* 32:349–371.

Rooth, Mats. 1992. A theory of focus interpretation. *Natural Language Semantics* 1:75-116.

Scarborough, Rebecca. 2010. Lexical and contextual predictability: Confluent effects on the production of vowels. *Laboratory Phonology* 10:557-586.

Selkirk, Elizabeth. 1984. *Phonology and Syntax: the Relation Between Sound and Structure*. Cambridge, Massachusetts: MIT Press.

Sharda, Megha, T. Padma Subhadra, Sanchita Sahay, Chetan Nagaraja, Latika Singh, Ramesh Mishra, Amit Sen, Nidhi Singhal, Donna Erickson, and Nandini C. Singh. 2010. Sounds of melody—Pitch patterns of speech in autism. *Neuroscience Letters* 478:42-45.

Shriberg, Elizabeth, D. Robert Ladd, Jacques Terken, and Andreas Stolcke. 1996. Modeling pitch range variation within and across speakers: Predicting f0 targets when "speaking up". In *Proceedings of the 4th international conference on spoken language processing*, 1-4.

Shue, Yen-Liang, Patricia Keating, Chad Vicenik, and Kristine Yu. 2011. VoiceSauce: A program for voice analysis. In *Proceedings of the 17th International Congress of Phonetic Sciences*, 1846–1849.

Smith, Rachel, and Sarah Hawkins. 2012. Production and perception of speaker-specific phonetic detail at word boundaries. *Journal of Phonetics* 40:213–233.

Theodore, Rachel M., Joanne L. Miller, and David DeSteno. 2007. The effect of speaking rate on voice-onset-time is talker-specific. In *Proceedings of the 16th International Congress of Phonetic Sciences*, 473-476.

Trouvain, Jürgen, and Martine Grice. 1999. The effect of tempo on prosodic structure. In *Proceedings of the 14th International Congress of Phonetic Sciences*, 1067–1070.

Vallduví, Enric, and Maria Vilkuna. 1998. On rheme and kontrast. In P. W. Culicover and L. McNally (Eds.), *Syntax and Semantics 29: The Limits of Syntax*. San Diego: Academic Press.

Van Donzel, Monique E., and Florien J. Koopmans-van Beinum. 1996. Pausing strategies in discourse in Dutch. In Proceeding of Fourth International Conference *on* Spoken Language *Processing (ICSLP '96)*, 1029–1032.

Van Son, R. J. J. H., Florien J. Koopmans-van Beinum, and Louis C. W. Pols. 1998. Efficiency as an organizing principle of natural speech. In *Proceedings of the 5th International Conference on Spoken Language Processing (ICSLP '98)*, 2375–2378.

Wagner, Michael, and Duane G. Watson. 2010. Experimental and theoretical advances in prosody: A review. *Language and Cognitive Processes* 25:905–945.

Ward, Gregory L., and Julia Hirschberg. 1985. Implicating uncertainty: The pragmatics of fall-rise intonation. *Language*, 747-776.

Ward, Gregory L., and Julia Hirschberg. .1986. Reconciling Uncertainty with Incredulity: A Unified Account of the L*+ HLH% Intonational Contour. Paper presented at the Annual Meeting of the Linguistic Society of America, New York, NY.

Watson, Duane G., Jennifer E. Arnold, and Michael K. Tanenhaus. 2008. Tic Tac TOE: Effects of predictability and importance on acoustic prominence in language production. *Cognition* 106:1548-1557.

Wennerstrom, Ann, and Andrew F. Siegel. 2003. Keeping the floor in multiparty conversations: Intonation, syntax, and pause. *Discourse Processes* 36:77-107.

Wright, Richard. 2003. Factors of lexical competition in vowel articulation. In John Local, Richard Ogden, and Rosalind Temple (Eds.), *Phonetic Interpretation: Papers in Laboratory Phonology VI*. Cambridge: Cambridge University Press.

Xu, Yi. 1999. Effects of tone and focus on the formation and alignment of F0 contours. *Journal of Phonetics* 27:55-105.

Yuan, Jiahong, Jason M.Brenier, and Dan Jurafsky. 2005. Pitch accent prediction: effects of genre and speaker. In *Proceedings of Interspeech 2005*, 1409-1412.

Appendix 1. Target items

The 48 critical sentences in the experiment are recoverable as follows. There are 12 conditions, formed by combining three types of question-response pairs (X-Z) and four kinds of object nouns in the responses (A-D). Each condition has four items, which can be differentiated based the verb-location context where the object nouns occurs (1-4). The subject of a question always consists of two personal names; no personal name occurs more than once in the experiment.

1. Context: *got…at the sports store*
   (X) Narrow Corrective Focus
   Partner asks: *I heard that {Dawn and Alice; …} got gloves at the sports store.*
   (Y) Narrow New-Information Focus
   Partner asks: *What did {Rachel and Carolyn; …} get at the sports store?*
   (Z) VP/Wide Focus
   Partner asks: *What did {Angela and Joyce; …} do?*
   (A) High Frequency and High Probability: *balls*
   (B) Low Frequency and High Probability: *cleats*
   (C) High Frequency and Low Probability: *fish*
   (D) Low Frequency and Low Probability: *toys*
   Participant responds: ***(No,) they got {balls; cleats; fish; toys} at the sports store.***

2. Context: *kicked…in the garage*
   (X) Narrow Corrective Focus
   Partner asks: *I heard that {Teresa and Martha; …} kicked dirt in the garage.*

(Y) Narrow New-Information Focus
Partner asks: *What did {Connie and Sharon; …} kick in the garage?*
(Z) VP/Wide Focus
Partner asks: *What did {Evelyn and Jacqueline; …} do?*
(A) High Frequency and High Probability: *cars*
(B) Low Frequency and High Probability: *cans*
(C) High Frequency and Low Probability: *books*
(D) Low Frequency and Low Probability: *shells*
Participant responds: ***(No,) they kicked {cars; cans; books; shells} in the garage.***

3. Context: *found…in the sea*
(X) Narrow Corrective Focus
Partner asks: *I heard that {Bonnie and Laura; …} found boats in the sea.*
(Y) Narrow New-Information Focus
Partner asks: *What did {Mary and Irene; …} find in the sea?*
(Z) VP/Wide Focus
Partner asks: *What did {Lillian and Gladys; …} do?*
(A) High Frequency and High Probability: *fish*
(B) Low Frequency and High Probability: *shells*
(C) High Frequency and Low Probability: *balls*
(D) Low Frequency and Low Probability: *cans*
Participant responds: ***(No,) they found {fish; shells; balls; cans} in the sea.***

4. Context: *found…on the stairs*
(X) Narrow Corrective Focus
Partner asks: *I heard that {Matthew and Edward; …} found socks on the stairs.*
(Y) Narrow New-Information Focus
Partner asks: *What did {Joseph and Steven; …} find on the stairs?*
(Z) VP/Wide Focus
Partner asks: *What did {Daniel and Jason; …} do?*
(A) High Frequency and High Probability: *books*
(B) Low Frequency and High Probability: *toys*
(C) High Frequency and Low Probability: *cars*
(D) Low Frequency and Low Probability: *cleats*
Participant responds: ***(No,) they found {books; toys; cars; cleats} on the stairs.***

Appendix 2. Lexical frequency of the target words

Word	Frequency in SUBTLEXus (per million)
fish	83.49

books	67.76
cars	45.63
balls	40.16
toys	13.22
cans	7.67
shells	5.57
cleats	0.41