Cognition 119 (2011) 58–69

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/COGNIT

Listeners' comprehension of uptalk in spontaneous speech

John M. Tomlinson Jr. [a,*], Jean E. Fox Tree [b]

[a] Department of Linguistics, Stanford University, United States
[b] Department of Psychology, University of California at Santa Cruz, United States

ARTICLE INFO

Article history:
Received 28 December 2009
Revised 1 December 2010
Accepted 6 December 2010
Available online 14 January 2011

Keywords:
Prosody
Intonational meaning
Pragmatics
Speech comprehension
Prolongations
Uptalk

ABSTRACT

Listeners' comprehension of phrase final rising pitch on declarative utterances, or *uptalk*, was examined to test the hypothesis that prolongations might differentiate conflicting functions of rising pitch. In Experiment 1 we found that listeners rated prolongations as indicating more speaker uncertainty, but that rising pitch was unrelated to ratings. In Experiment 2 we found that prolongations interacted with rising pitch when listeners monitored for words in the subsequent utterance. Words preceded by prolonged uptalk were monitored faster than words preceded by non-prolonged uptalk. In Experiment 3 we found that the interaction between rising pitch and prolongations depended on listeners' beliefs about speakers' mental states. Results support the theory that temporal and situational context are important in determining intonational meaning.

1. Introduction

In the following exchange, Stephanie makes a blanket statement about violence in video games. What is unusual about this declarative utterance is that it ends with something more indicative of interrogative utterances: rising pitch (∗ = pitch accent, / = rising pitch, commas indicate brief pauses):

Stephanie: I don't think ∗ that video games ∗ make you become violent/
Marcus: Uh-huh
Stephanie: Yeah, because people ∗ can separate reality ∗ from, uh, like, video games.

Two reasons why Stephanie might raise her pitch at the end of her stated opinion are that: (1) she is indicating that she is unsure of the truth propositional content of her utterance, either whether she believes it or whether Marcus believes it, or (2) she believes what she is saying but has yet to provide further information bolstering her claim. These two functions can be differentiated as to whether they apply to the utterance she just stated ("I don't think that video games make you become violent"), a *backward-looking* function, or to the subsequent utterance ("yeah, because people can separate reality from uh like video games"), a *forward-looking* function (Allen, 1984; Pierrehumbert & Hirschberg, 1990; Fletcher, Stirling, Mushin, & Wales, 2002; House, 2007). These competing functions create an intonational ambiguity for the listener. Similar to syntactical and lexical ambiguities, intonational ambiguities help elucidate how intonation is interpreted.

Rising declarative pitch, or *uptalk*, is sometimes considered a restricted dialectical variation of American English (McLemore, 1991).[1] In fact, however, uptalk is common in Australian English (Allen, 1984), Southern British and Belfast English (Cruttenden, 1997), and Canadian English (Shokier,

∗ Corresponding author. Address: Linguistics and Symbolic Systems, Margaret Jacks Hall, Bldg. 460, Stanford University, Stanford, CA 94305-2150, United States. Tel.: +1 650 725 1562; fax: +1 650 723 5666.
E-mail addresses: tomlinsonj2@cardiff.ac.uk (J.M. Tomlinson), foxtree@ucsc.edu (J.E. Fox Tree).

0010-0277/$ - see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.cognition.2010.12.005

[1] The term *uptalk* was used here because it encompasses rising pitch starting from both the lower and upper parts of a speaker's pitch range, not just the upper part as historically understood with the label *high rise terminals* (Liberman, 2006).

2008). Despite the ubiquity of uptalk in multiple dialects of English, it has been cast aside by many theoreticians as an anomaly or as meaningless sociolinguistic variation because of the assumption that all declarative utterances in English end with falling pitch, or declination. Declination may be a product of scripted talk, however, with possibly little or no declination in spontaneous talk (Ohala, Dunn, & Sprouse, 2004). Also, there seems to be no difference between how older and younger adults use uptalk versus falling pitch, which suggests that uptalk is not an ongoing linguistic shift (Shokier, 2008), although this result merits replication with a larger sample size. From this we conclude that uptalk is a productive linguistic phenomenon.

We propose that listeners disambiguate the proposed two functions of uptalk using: (1) the temporal context of prolongations and (2) expectations about the speakers' knowledge states. In Experiments 1a and 1b, we tested listeners' off-line interpretations of prolongations and uptalk with respect to listeners' perceptions of speaker accuracy and certainty. In Experiment 2, we tested the speed at which listeners monitored for words following prolongations or uptalk. In Experiment 3, we tested how listeners' beliefs about speakers' knowledge states influenced their on-line word monitoring performances.

2. Background

Much work on intonational meaning has concentrated on mapping intonational form to *information structure* (Chafe, 1976), where the information status of a concept changes based on *mutual belief space* (Pierrehumbert & Hirschberg, 1990). For example, a speaker can produce a specific pitch accent or tune to update the mutual belief space shared with listeners. For uptalk, Pierrehumbert and Hirschberg argued the following (1990, p. 288):

> We propose that tune meaning is composed of the meanings of three types of tone – pitch accents, phrase accents, and boundary tones – which have scope over three different domains of interpretation. Together, these intonational features can convey how S[peaker] intends that H[earer] interpret an intonational phrase with respect to (1) what H already believes to be mutually believed [backward-looking] or (2) what S intends to make mutually believed as a result of the subsequent utterances [forward-looking].

In essence, the Pierrehumbert and Hirschberg (1990) approach argues that a given tone or tune has a unique information-theoretic function. In this section, we will review empirical research that has tested this approach. Also, we explore the idea that some of the functions resulting from intonational forms go above and beyond information theoretic accounts of intonational meaning.

Abundant empirical evidence for this approach to intonational meaning can be found in off-line studies of corpora, map tasks, and listener judgments (Beckman & Pierrehumbert, 1986; Hirschberg & Pierrehumbert, 1986; Hirschberg & Ward, 1992; Kowtko, 1996; Ladd & Morton, 1997). Some on-line studies also support one-to-one intonational form-function relationships. For example, off-line studies of contrastive pitch accents have shown that *upsteps*, or L + H* accents, generally mark contrastive information (Bartels & Kingston, 1994; Cutler, 1976; Cutler & Clifton, 1984; Guhe, Steedman, Bard, & Louwerse, 2006; Krahmer & Swerts, 2001), and on-line studies support this interpretation. When accent was used on adjectives in an interactive Christmas Tree decoration task ("Hang the GREEN ball above the blue ball"), listeners demonstrated anticipatory eye movements to the ball of the contrastive color (the blue ball; Ito & Speer, 2008).

But other on-line work has offered a different understanding of intonational form-function relationship – one in which a single intonational form can have multiple, and sometimes overlapping, functions. For example, an initial claim for the *given-new contract* (Clark & Bangerter, 2004; Clark & Haviland, 1977; Garrod & Sanford, 1982) was that accented syllables (H*) uniquely mark new additions to mutual belief space, whereas deaccented syllables (L) mark given information (see Pierrehumbert & Hirschberg, 1990, for how L could be used in other ways; see Prince, 1981, for further discussion of given information). Many off-line studies have provided support for these claims (Birch & Clifton, 1995; Bock & Mazzella, 1983; Terken & Nooteboom, 1987). But on-line measures do not necessarily support this approach. Dahan, Tanenhaus, and Chambers (2002) found that listeners have more predictive looks to unmentioned referents when the first syllable of two lexical competitors (either *candy* or *candle*) is accented (H*). However, in their second experiment, in which the target (*candy* or *candle*) was mentioned as the goal of the previous sentence, this time with unaccented first syllables, listeners fixated the already mentioned word (*candle*) upon hearing the accented syllable (CAN). Critically, this shows that pitch accenting can comprise multiple functions, including marking discourse information as new and making given information salient. Furthermore, this finding supports previous research from speech production suggesting that the relationship between given information and intonation might be better understood through *accessibility* as opposed to a purely referential account (Bard & Aylett, 1999; Baumann & Grice, 2006; Watson, Gunlogson, & Tanenhaus, 2006).

On-line work also changed understanding of H* versus L + H* accents as markers of contrast. In an eye-tracking study, L + H* accents seemed to uniquely signal contrast, but H* accents signaled both contrast and new information (Watson, Tanenhaus, & Gunlogson, 2008). This suggests that instead of simply marking new information as proposed by Pierrehumbert and Hirschberg (1990), the H* accent might serve to increase the salience of previously less salient information. In sum, these on-line studies do not necessarily support the Pierrehumbert and Hirschberg (1990) approach to intonational meaning because an individual tone (H*) seems to entail two overlapping functions instead of a strict one-to-one mapping between intonational form and function. As a consequence, if an intonational form contains multiple functions, how might listeners go about differentiating one function from the other? Next, we review evidence that suggests that uptalk might also have multiple functions, and propose several hypotheses about what type of information might

differentiate these functions and how listeners use that information.

To date, no on-line studies have been conducted to test the distinction between backward-looking and forward-looking functions of uptalk. But there is reason to expect that on-line results for uptalk might differ from those identified for the given-new distinction and contrastive stress. Both given-new and contrastive stress affect the information status of a constituent. But uptalk coordinates interlocutors' beliefs about their shared knowledge (Clark, 1996; Pickering & Garrod, 2004). This distinction between *common ground content*, which refers to temporary or static belief state of the addressee's mental state at the point of production, and *common ground management*, which represents the non-factual informational needs of interlocutors to update common ground in real time, may yield different on-line effects (Krifka, 2007).

There is some off-line work about the backward-looking function, all of it supporting the hypothesis that rising pitch signals uncertainty. When question answerers were unsure of their answers to factual questions, they were more likely to produce answers with uptalk (Smith & Clark, 1993). Correspondingly, listeners thought speakers were less likely to know answers when answers were produced with delayed uptalk (Brennan & Williams, 1995). In another study, listeners inferred that a referent mentioned with rising pitch was atypical (Barr, 2003). Utterances with yes/no question intonation (L*H–H%) were rated as highly uncertain, and those with *downsteps* (!H*L–L%) were rated as more certain relative to simple declarative contours (Gravano, Benus, Hirschberg, Sneed, & Ward, 2008).

There is also some off-line work about the forward-looking function. Uptalk in rising declaratives has been described as similar to list intonation, which differs from yes/no intonation by having a low range rise preceded by either an accented or unaccented syllable (H*L–H% or L*L–H%; Shokier, 2008). While yes/no intonation might reasonably be associated with uncertainty, rising declaratives might be associated with certainty. In a corpus analysis of Australian English, low onset high rises (L*H–H%) were more associated with forward-looking functions, such as continuation, than with yes/no requests (Fletcher et al., 2002). In contrast, low range rises, or list intonation (H*L–H%), were found to be more backward-looking in that they corresponded to requests for information (Fletcher et al., 2002). In another study, H–H% contours were judged more likely to indicate continuity than L–H% contours, although they were also rated as less certain than L–H% (Shokier, 2008).

Taken together, the off-line studies demonstrate that listeners do not reliably associate L–H% or H–H% tones with particular functions. That is, the nature of the tune itself might not be sensitive enough to distinguish the forward-looking and backward-looking functions of rising pitch.

How might listeners differentiate the two functions of uptalk? We propose that time may be the missing ingredient. That is, in addition to the intonation of an utterance, how an utterance unfolds temporally may contribute to the meanings listeners construe. This hypothesis runs contrary to existing theories of intonational meaning. For example, Pierrehumbert and Hirschberg (1990) claim that "[a]lthough the interpretation of any token of a tune type may vary along many other dimensions – voice quality, pitch range as well as non-intonational features – any instance of a given tune will convey the same relationship" (1990, p. 285). But timing can affect functional interpretations. Because varying the temporal alignment of pitch accents can alter listeners' interpretations of function (Kohler, 2006), how uptalk unfolds temporally might play an important role in how listeners go about distinguishing the conflicting functions of uptalk.

We argue, as do others, that intonation is inherently temporal (Clark, 2002; Kohler, 2004; Ramus & Mehler, 1999). As a consequence, how intonation is expressed temporally might affect its interpretation. Because uptalk occurs in a phrase final position, it may be subject to phrase final lengthening (Selkirk, 1995). The final syllable of an utterance can be prolonged as a result of syntactic or prosodic structure (Ferreira, 1993; Gee & Grosjean, 1983; Watson & Gibson, 2004), conversational pressures (Anderson et al., 1991; Ward & Tsukahara, 2000; Clark, 2002), or the dynamic modulation of articulators (Byrd & Saltzman, 2003).

3. Uptalk and prolongations

How might listeners process upcoming information upon hearing a prolongation and how might prolongations influence how uptalk is interpreted? We will first review evidence about prolongations, and then propose how prolongations may interact with uptalk.

Prolongations can both indicate that a delay is in progress and predict upcoming delays (Clark & Fox Tree, 2002; Fox Tree & Clark, 1997). For example, pauses, fillers, and repairs were about five times as common surrounding the word *a* pronounced *ay* than pronounced *uh* (Fox Tree & Clark, 1997). Sometimes the delays are more common after prolongations. Pauses and fillers were about twice as common before the word *the* when it was pronounced as *thiy* (or *thee*) than when it was pronounced as *thuh*, but they were nineteen times as common after *thiy* than *thuh* (Fox Tree & Clark, 1997). This pattern is similar when looking at only the pronunciation *thiy* with a shorter and longer vowel; *thiy* and *thi:y* were preceded by pauses and fillers at about the same rate, but *thi:y* was followed by pauses and fillers at almost 1.5 times the rate (Fox Tree & Clark, 1997). In contrast, when comparing prolonged (*u:m* and *u:h*) and non-prolonged (*um* and *uh*) fillers, there were only slightly more and longer delays following prolonged fillers than non-prolonged fillers, but there were almost twice as many delays before prolonged fillers, and those delays were twice as long (Clark & Fox Tree, 2002).

Because delays after *um* and *u:m* were twice as common as delays after *uh* and *u:h*, Clark and Fox Tree (2002) proposed that the choice between *um* versus *uh* and between prolonged and non-prolonged were independent signals of delay, with prolongations responding to prior delay and *um* versus *uh* signaling upcoming delay. Other research suggests that the choice between prolongation and filler may depend on the language speakers speak (Dunn & Fox

Tree, 2009; Fox Tree, 2010). Spanish-weighted speakers prefer prolongation, and English-weighted speakers prefer fillers (Dunn & Fox Tree, 2009).

But regardless of whether there are more delays before or after prolonged words, and whether this depends on the word prolonged (*um* versus *the*) or the language spoken (Spanish or English), prolongations are strongly associated with delays. Recognizing delays can focus listeners' attention on upcoming talk, but only if those delays are anticipated to be short (Fox Tree, 2001). Therefore, we predict that prolongations have an effect on the processing of subsequent speech only if listeners interpret them as indicating a brief upcoming delay.

In the absence of delay, uptalk applied to an utterance suggests that the speaker is not taking extra time contemplating the currently produced utterance (prolongations indicating a delay in progress) or the upcoming utterance (prolongations indicating an upcoming delay). This suggests that the current utterance is fully formed. We propose that in the absence of evidence to the contrary, the resulting fully-formed declarative sentences produced with questioning intonation are interpreted as either (1) solicitations of acknowledgement of grounding of that utterance, alerting listeners that they should attest to the relevance of the utterance, or (2) shifting the responsibility to listeners to determine an utterance's truth value; that is, the non-prolonged uptalk expresses uptalk's backward-looking function. We say "in the absence of evidence to the contrary" because it is possible that under some circumstances, non-prolonged uptalk can be interpreted differently by listeners. For example, if listeners believe the speaker they are overhearing is talking to a knowledgeable addressee, the uptalk may be interpreted as solicitations of acknowledgement of grounding from that knowledgeable addressee, but not of the overhearers themselves, allowing the overhearers to distance themselves from the effects of the prolongations and uptalk.

In the presence of delay, uptalk applied to an utterance suggests that the speaker is taking extra time contemplating either the currently produced utterance or the upcoming utterance. We propose that in the absence of evidence to the contrary, the resulting declarative sentence with prolonged uptalk will suggest, at least on some occasions, that the upcoming utterance is requiring extra effort and therefore may be worthy of extra attention from the listeners. Because prolongations can focus attention on upcoming information, we propose that uptalk in combination with prolongations suggests to listeners the forward-looking function of uptalk, which is to focus on what's about to be said to link it to the prior information. Once again, we say "in the absence of evidence to the contrary" because it is possible that under some circumstances, uptalk can be interpreted differently. If listeners believe the speaker they are overhearing is talking to a knowledgeable addressee, the prolonged uptalk directing attention at upcoming utterances may be interpreted as uninformative for the overhearer.

Taken together, we propose that if uptalk is produced without prolongation, listeners will associate the non-prolonged uptalk with the backward-looking function, directing attention to the utterance produced with the uptalk and away from the upcoming utterance. If the uptalk is produced with prolongation, listeners will associate the prolonged uptalk with the forward-looking function, directing attention away from the utterance produced with the uptalk and towards the upcoming utterance.

There is some on-line work with respect to forward-looking functions. In a semantic verification study, listeners responded quicker to a visual item after they heard it stressed in an utterance than when the item was not stressed (Gernsbacher & Jescheniak, 1995). This demonstrated that pitch accenting can signal to listeners that a concept should remain accessible in their discourse representation because it is likely to be referred to again. In their conclusion, Gernsbacher and Jescheniak (1995) suggested that phrase final rising pitch may also serve a cataphoric function.

There is also some precedent for listeners' inferences affecting on-line processing. Information about a speaker's ability to produce the names of words influenced on-line responses to what the speaker said (Arnold, Hudson Kam, & Tanenhaus, 2007). If the speaker were considered a normal language user, hearing a disfluency prompted listeners to look at previously unmentioned items in an array of items. But if the speaker were thought to have object agnosia, listeners did not necessarily look at the unmentioned items. Similarly, intonational events might evoke rapid inferences about speakers' communicative intentions and as a result help coordinate listeners' attention to upcoming information. Unlike research assessing the effect of pitch accents on local referential expressions, this study presents a novel approach by examining more global aspects of intonational meaning, which need not be anchored to a specific referent. Moreover, we seek to shed light on how the temporal realization of uptalk (prolongations) might effect listeners' on-line comprehension and disambiguation of the two conflicting functions of uptalk (forward vs. backward-looking).

In three experiments, we tested how listeners establish relationships between intonational form and function. In Experiment 1, we tested the hypothesis that rising pitch signals knowledgeableness and speaker certainty to the listener, relative to falling pitch. Studies on the *feeling of knowing*, defined here as speakers' and listeners' assessments of their conversational partner's level of knowledgeableness during production, predict that rising pitch should be interpreted as less knowledgeable than falling pitch (Barr, 2001; Brennan & Williams, 1995; Smith & Clark, 1993). Feeling of knowing studies do not provide a prediction for prolongations.

In Experiment 2, we tested the effect of rising pitch and prolongations on the monitoring of upcoming information. When listeners expect a brief delay after hearing an *uh*, attention is focused on upcoming talk and word monitoring is faster (Fox Tree, 2001). Prolongations can mark ongoing delays, anticipating briefer delays after them than before them (Clark & Fox Tree, 2002). Similar to *uh*s (Fox Tree, 2001), listeners may focus on upcoming words after prolongations; that is, prolongations may direct listeners' attention forward rather than backward. While earlier studies make no prediction for the effect of pitch on word monitoring, phoneme-monitoring studies show that pitch

events are sensitive enough to be used to predict information in the upcoming utterance (Cutler, 1976).

In Experiment 3, we tested the role of listener expectations on on-line speech comprehension. Half the participants were told that speakers had memorized facts from cue cards and had trouble reproducing them later because they were not experts in the relevant domains. Half were told that the speakers were experts in a domain related to the content of what they were saying. If listeners interpret rising pitch and prolongations with a predetermined set of rules, then reaction times should be similar between the expert and non-expert conditions. But if listeners take information about speakers into account when listening to talk, reaction times might differ. For example, if the speakers are thought to be non-experts, then rising pitch may serve a backward-looking function yielding slower reaction times. But if speakers are thought to be experts, rising pitch may serve a forward-looking function yielding faster reaction times.

4. Experiments 1A and 1B

In Experiment 1A, we tested whether listeners attributed different levels of speaker *accuracy* to rising pitch and prolongations. In Experiment 1B, we tested whether listeners attributed different levels of *certainty* to rising pitch and prolongations. Listeners rated utterance pairs on how accurate and certain speakers were in the knowledge they were conveying.

4.1. Methods

4.1.1. Participants

In exchange for course credit, 20 University of California Santa Cruz (UCSC) students participated in Experiments 1A and 13 participated in Experiment 1B. All were native speakers of English.

4.1.2. Materials

Forty-eight utterance pairs were selected from a specially compiled corpus of spontaneous speech. Speakers created spontaneous sentence frames to convey celebrity facts to an addressee who attempted to select the celebrity out of an array. For example, upon reading *place of birth*: *Brooklyn*, the speaker might say to the addressee, "This actor was born in Brooklyn, New York." The utterance pairs consisted of two sentences spontaneously produced by the same speaker; for example, "I have two children. I was the princess of Wales."

The first utterance of the pair had either rising prolonged pitch ($N = 12$), rising non-prolonged pitch ($N = 12$), falling prolonged pitch ($N = 12$), or falling non-prolonged pitch ($N = 12$). Although many different transcriptional schemes could be used to describe our stimuli, we chose to use ToBI, *Tone and Break Indices* guidelines (Beckman & Ayers, 1994), because we anticipate that many readers are familiar with it. Our corpus contained utterances with two types of rising pitch and three types of falling pitch. Rising pitch consisted of either H*L–H% (a pitch accented syllable followed by a low range rise) or L*H–H% (a low accented syllable followed by a high range rise), whereas falling pitch consisted of either abrupt falls such as H*L–L% (a pitch accented syllable followed by high range fall) or more level falls such as H*H–L% (a pitch accented syllable followed by more level fall) or L*H–L% (a low accented syllable followed by more level fall).

Prolonged and non-prolonged syllables were primarily classified quantitatively: non-prolonged syllables averaged 291 ms (SD = 133 ms; range 228 ms–433 ms), whereas prolonged syllables averaged 501 ms (SD = 156 ms; range 313 ms– ms). Three prolonged rises fell in the non-prolonged range but were classified as prolonged because of the nature of the rising pitch contour. For example, the final "C" in "I was born in Washington DC" only has a duration of 403 ms; however, the rising contour reaches a ceiling towards the middle of the segment and is further drawn out across the rest of the segment (see Fig. 1). All of the non-prolonged rises contained rising pitch contours that reached ceiling (F0 maxima) towards the end of the segment.

For each utterance with a rising contour, a matched utterance was digitally stylized with a falling contour using the PSOLA function on PRAAT (Boersma & Weenink, 2008). Likewise, for each utterance with a falling contour, a matched utterance was digitally created with a rising contour. The beginning of the rise or fall was kept the same. The ascension or descension of the contour was multiplied or divided by a factor of 1.5 and then further stylized to smooth out the contour (see Fig. 2). For example, a rising contour starting in the higher part of a speaker's pitch range would become a falling contour that started in the higher part of the speaker's pitch range. In this way, the same item could be heard with and without uptalk. Prolongation was treated as a between-item variable.

4.1.3. Design

Two lists were created so that each participant heard either the original or the manipulated version of an item. The lists were counterbalanced so that: (1) original and manipulated versions were matched across lists, (2) half the items on each list were manipulated and half were not, and (3) rises and falls were equally likely to occur on a list.

4.1.4. Procedure

Participants were told that the speakers they would listen to were recalling facts they had learned about celebrities, and that the speakers may have muddled their facts. In Experiment 1A, participants judged the likelihood that the speaker correctly reported the facts. The items were presented aurally with a screen displaying a 1 (not accurate) to 7 (accurate) Likert scale. In Experiment 1B, participants judged the likelihood that the speaker was certain about the facts. The items were presented aurally with a screen displaying a 1 (not certain) to 7 (certain) Likert scale.

4.2. Results

Here and in the subsequent two experiments, a *mixed-effects model* was conducted to test the effects and

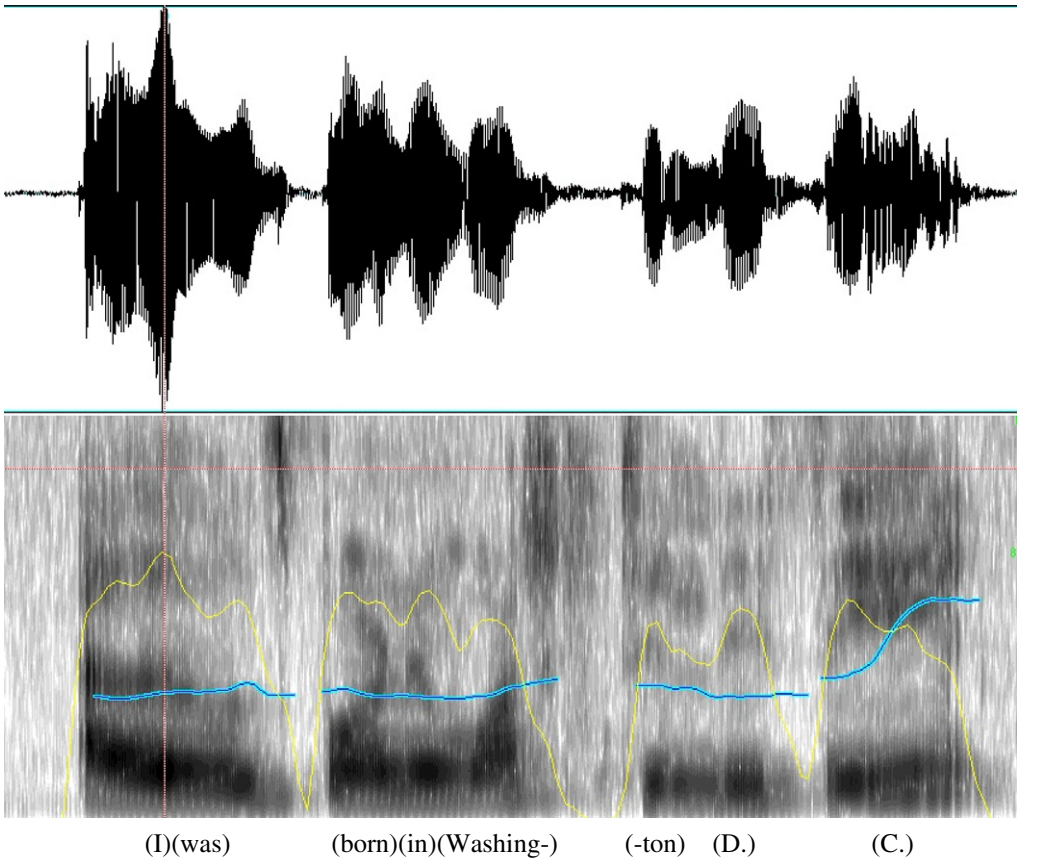

Fig. 1. Prolonged rise – "I was born in Washington, DC".

interaction of rising pitch and prolongations using the *lmer* (Bates, 2007) and *languageR* (Baayen, 2007) packages in R (R Development Core Team, 2007). Linear mixed-effects regression models were calculated with both subjects and items as crossed random effects and the *p*-values were obtained using Markov Chain Monte Carlo (MCMC) sampling (Baayen, Davidson, & Bates, 2008; Quené & van den Bergh, 2008). Because items were matched for rising and non-rising pitch, however not matched for prolongations, the model included random intercepts for items and subjects. A Nagelkerke pseudo r^2 was calculated for the entire model as a measure of effect size for each experiment.[2]

In Experiment 1A, listeners rated speakers as less accurate when speakers prolonged syllables at the end of the first utterance in the pair, $t = 4.31$, $p < .01$. Accuracy ratings were unaffected by rising pitch at the end of the first utterance in the pair, $t = 1.2$, $p = .23$. There was no interaction, $t = 1.76$, $p = .18$. Accuracy ratings were similar across edited and unedited items, $t(47) = .82$, $p = .42$. For the model, the Nagelkerke pseudo $r^2 = .15$. See Table 1.

In Experiment 1B, listeners rated speakers as less certain when speakers prolonged syllables at the end of the first utterance in the pair, $t = 3.56$, $p < .01$. Certainty ratings were unaffected by rising pitch at the end of the first utterance in the pair, $t = 0.06$, $p = .961$. There was no interaction, $t = 0.98$, $p = .38$. Certainty ratings were similar across edited and unedited items, $t(47) = 0.98$, $p = .34$. For the model, the Nagelkerke pseudo $r^2 = .25$. See Table 2.

4.3. Discussion

Prolongations decreased ratings of speaker knowledgeableness as assessed by off-line ratings of accuracy and certainty. Although earlier studies predicted rising pitch would indicate lack of knowledgeableness, we found no evidence for this. Earlier tests may have found a relationship because judgments were based on answers to factual questions. Listeners may construe uptalk differently when judging answers to factual questions. Judgments in our study were of the first statement in a pair of statements. These statement pairs, although realistically modeling

[2] There is currently no agreement over a standard measure of goodness of fit for logit mixed-effect models. The Nagelkerke pseudo r^2 was chosen because its value falls between 0 and 1, hence having the same scale as a traditional r^2. This statistic should be interpreted with caution however due to lack of generalization across data sets. The Nagelkerke pseudo r^2 was computed by dividing the log likelihood of the model by the log likelihood of the intercept, squaring it by $2/N$, and then subtracting it from 1.

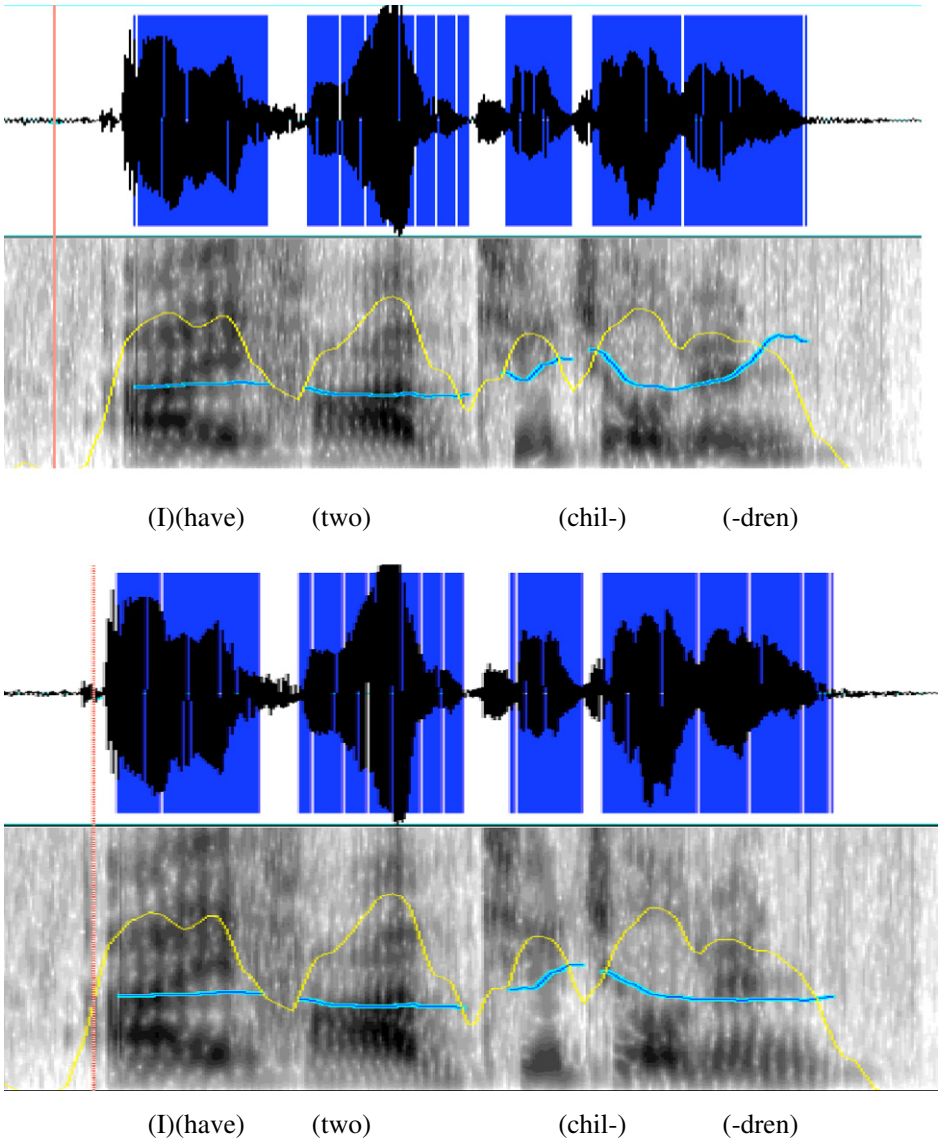

Fig. 2. Non-prolonged rise (top) to non-prolonged fall (bottom) – "I have two children".

Table 1
Experiment 1A mean ratings (SD) from 1 (not accurate) to 7 (accurate).

	Prolonged	Non-prolonged
Fall (H*L–L%, L*H–L%, or H*H–L%)	4.28 (1.07)	4.96 (.76)
Rise (L*H–H% or H*L–H%)	4.44 (.87)	4.91 (.88)

Table 2
Experiment 1B mean ratings (SD) from 1 (not certain) to 7 (certain).

	Prolonged	Non-prolonged
Fall (H*L–L%, L*H–L%, or H*H–L%)	4.08 (.94)	4.96 (.74)
Rise (L*H–H% or H*L–H%)	4.43 (.82)	4.99 (.91)

uptalk in natural dialogue, do afford a less obvious connection between the utterance and speaker knowledgeableness than answers to factual questions.

Another reason why these off-line ratings may not have yielded an effect of rising pitch is that listeners may have been unsure how to interpret the two hypothetical functions of rising pitch in this task. Participants in Experiments 1A and 1B may not have been able to judge, off-line, whether the speakers they listened to were engaging in backward-looking fact checking or forward-looking directing of attention. However, listeners may be prompted to behave in different ways when faced with an on-line task. When monitoring for words in the utterance following the manipulated utterances, listeners may respond more

quickly or slowly depending on whether the prior utterance ended with prolongations or uptalk or both.

5. Experiment 2

In Experiment 2, we tested whether prolongations influenced whether uptalk was interpreted with forward-looking or backward-looking functions. Listeners monitored for words that followed a sentence that ended with either a prolongation, uptalk, both a prolongation and uptalk, or neither a prolongation nor uptalk. We predict that listeners will monitor for words faster after prolonged uptalk, which we predict will focus attention on upcoming information, than non-prolonged uptalk, which we predict will not focus attention on upcoming information, as it has a backward-looking function.

5.1. Methods

5.1.1. Participants

In exchange for course credit, 21 UCSC students participated. All were native speakers of English.

5.1.2. Materials

The same 48 stimuli were used as in Experiment 1. In addition, 48 filler stimuli were created. The filler stimuli were of two types. The first had the target word in the first utterance. The second did not contain the target word. When no target word was presented in the stimulus, there was a 1000 ms pause until the next trial started.

5.1.3. Design

The same design as in Experiment 1 was used, with the addition of filler stimuli.

5.1.4. Procedure

Each trial had the following structure. First, a 500 ms tone was heard indicating that the participants should focus their attention on the computer screen. The tone was followed by a 500 ms pause. A word appeared on the computer screen for 1000 ms, followed by a blank screen for 1000 ms, followed by the onset of the auditory stimulus. During the auditory stimulus, listeners pressed a button upon hearing the word that was previously presented visually on the screen. All target items and half of the fillers required a button press. The other half of the fillers required no button press. Targets were in variable positions in the second utterances. Half of the fillers contained targets in the first utterances and the rest of the targets contained no targets at all. When no target word was presented in the stimulus, there was a 1000 ms pause until the next trial started.

5.2. Results

One participant was removed for having average reaction times over 3000 ms in each of the four categories. Response times more than two standard deviations from the mean were treated as outliers and were removed from the analysis (7% of the data). This eliminated both false alarms and misses.

Table 3
Experiment 2 mean reaction times (SD) in milliseconds.

	Prolonged	Non-prolonged
Fall (H*L–L%, L*H–L%, or H*H–L%)	545 (90)	532 (85)
Rise (L*H–H% or H*L–H%)	517 (83)	589 (134)

Listeners tended to monitor words faster after hearing prolonged syllables in the previous utterance, however this effect was only marginally significant, $t = 1.679$, $p = .09$. There was no effect of rising pitch, $t = 0.195$, $p = .84$, but there was an interaction between prolongations and rising pitch, $t = 2.66$, $p < .001$. Listeners monitored words fastest after prolonged rises and slowest after non-prolonged rises. There was no effect of token manipulation on reaction times, $t(47) = 0.67$, $p = .51$. For the model, the Nagelkerke pseudo $r^2 = .41$. See Table 3.

5.3. Discussion

Words were monitored for fastest when they followed sentences ending in prolonged rising pitch. Words were monitored for slowest when they followed sentences ending in non-prolonged rising pitch. These data show that prolongations can differentiate the two conflicting functions of rising pitch. Listeners tended to gain a general processing advantage when comprehending material following prolongations. At the same time, prolongations interacted with rising pitch in a way that makes prolonged uptalk more helpful for processing upcoming information and non-prolonged uptalk more harmful for processing upcoming information.

In a model with predetermined form-function relationships, these data would suggest that prolonged rising pitch equals forward-looking, and non-prolonged rising pitch equals backward-looking. But it may be the case that listeners use beliefs about speakers' mental states in mapping intonational forms to different functions. In Experiment 3, we test how listeners' beliefs about the speaker affect processing.

6. Experiment 3

Listeners replicated Experiment 2 with new information about the speakers they would be hearing. Half were told that the speakers were non-experts who had memorized the facts off cue cards and had trouble remembering them during the game. The other half were told that the speakers were knowledgeable experts on the facts they were saying.

6.1. Methods

6.1.1. Participants

In exchange for course credit, 48 UCSC students participated. All were native speakers of English.

6.1.2. Materials

The same materials were used as in Experiment 2.

6.1.3. Design

The design was similar to Experiments 1 and 2, except that Experiment 3 was effectively two separate experiments. In one, 24 participants were told that the speakers were non-experts in the fields they were talking about; that is, they knew little about pop culture, cinema, politics, and so on. In the other, 24 participants were told that the speakers were experts in the fields they were talking about. For example, they were led to believe that a person talking about politicians was a political science major and that a person talking about actors was a film major.

6.1.4. Procedure

The procedure was the same as in Experiment 2.

6.2. Results

Two participants were removed for having average reaction times over 3000 ms in each of the four categories. Response times more than two standard deviations from the mean were treated as outliers and were removed from the analysis (5% of the data). This eliminated both false alarms and misses.

As in Experiment 2, reaction times were calculated by subtracting the target onset from the overall reaction time. Overall, the results replicated the interaction between pitch and prolongation from Experiment 2, $t = 3.04$, $p < .01$, however only for the non-expert condition and not for the expert condition. There was a 3-way interaction between expertise, pitch, and prolongation, $t = 2.97$, $p < .01$. There was no main effect of prolongation, $t = 1.52$, $p = .13$, and no main effect of pitch, $t = 0.73$, $p = .48$. Last, there was no effect of token manipulation on reaction times, $t(47) = .17$, $p = .97$. For the model, the Nagelkerke pseudo $r^2 = .55$ See Table 4.

6.3. Discussion

When listeners believed that speakers were non-experts, they interpreted rising pitch and prolongations similarly to when they were provided with no information about speakers (Experiment 2). However, when listeners thought that speakers were experts, their word monitoring was no longer affected by rising pitch or prolongations. This is strong evidence that listeners establish relationships between linguistic form and function by first presupposing speakers' mental states.

Table 4
Experiment 3 mean reaction times (SD) in milliseconds.

	Prolonged	Non-prolonged
Non-expert condition		
Fall (H*L–L%, L*H–L%, or H*H–L%)	556 (130)	544 (122)
Rise (L*H–H% or H*L–H%)	515 (109)	624 (155)
Expert condition		
Fall (H*L–L%, L*H–L%, or H*H–L%)	572(172)	583 (170)
Rise (L*H–H% or H*L–H%)	564 (130)	574 (189)

7. General discussion

The temporal realization of intonational events affected listeners' interpretations of those events, in both off-line and on-line tasks. At the same time, listeners' expectations also influenced how intonational events were interpreted on-line. These findings necessitate a reexamination of mainstream theories of intonational meaning.

In Experiment 1, listeners considered utterances with prolongations as lacking in knowledgeableness and certainty. In contrast to expectation, uptalk had no effect on listener judgments. It may be that the off-line rating task was not sensitive enough to capture how listeners interpreted rising pitch with the stimuli we used. In earlier studies, rising pitch indicated a lower feeling of knowing with respect to the answers to questions. In our study, rising pitch was used with the first declarative statement in a pair of statements. With these materials, there could be two interpretations for rising pitch, forward-looking and backward-looking. Also, it might be that off-line experiments allow listeners to engage in inferential processes that might not be possible, or at least highly constrained, in an on-line study.

In Experiment 2, although prolongations tended to give listeners a processing advantage for monitoring words in the subsequent utterance, the only reliable effect was an interaction between prolongations and uptalk. Prolonged rising pitch engendered the fastest reaction times for monitoring words in the subsequent utterance. Non-prolonged rising pitch engendered the slowest reaction times for monitoring words in the subsequent utterance. This supports the hypothesis that prolonged rising pitch is forward-looking, and that non-prolonged rising pitch is backward-looking. Rising pitch alone had no effect.

In Experiment 3, the interaction between rising pitch and prolongations was replicated for the group of listeners who believed that the speakers they were hearing were unfamiliar with the material they were talking about. The interaction disappeared for another group of participants who believed that the speakers were experts on what they were talking about.

Taken together, these findings demonstrate that intonational events can be interpreted differently depending on temporal and situational context. What does this mean for mainstream theories of intonational meaning? There are at least two potential consequences of our findings.

First, these findings demonstrate that temporal context can influence the meaning of intonational events. In the autosegmental approach and its associated transcriptional scheme ToBI (Beckman & Ayers, 1994), the timing of intonational events are only relevant in the alignment of intonational events to prosodic phrasing; they do not play an active role in the interpretation of the intonational event itself. Pitch accented syllables usually display other prosodic characteristics such as higher amplitude and longer duration (see Hirschberg & Ward, 1992), and therefore our results might simply be explained as driven by a pitch accented syllable followed by a rise. Indeed, the alignment of pitch accents to syllables (early peak vs. late peak rises) has been shown to drastically affect how listeners interpret

intonational meaning (Kohler, 2006). Although we agree that pitch accenting and duration have a strong relationship, this account cannot explain the cross-over interaction between prolonged versus non-prolonged final syllables and rises: Why would a non-prolonged rise slow down processing of upcoming information?

Second, results from Experiment 3 reinforce the idea that common ground, or mutual belief space (Pierrehumbert & Hirschberg, 1990; Clark, 1996), plays an essential role in determining intonational form-function relationships. This also supports recent research on the spontaneous production of rising declaratives and how they are used to coordinate joint attention as a function of a speaker's belief about shared (visual) common ground (Richardson, Dale, & Tomlinson, 2009).

To remedy the limitations of the autosegmental account, we will now briefly consider an alternative inference-based account that might better explain how listeners establish the meaning of uptalk. As mentioned in the introduction, the autosegmental account has had success explaining data from off-line studies; however, it has not provided the nuanced approach needed to explain some recent on-line studies. Whereas the autosegmental account seems to suggest that intonational meaning consists of a grammatical relationship between intonational form and function, the inference-based account of intonational meaning emphasizes conversational implicature in determining intonational form-function relationships (Grice, 1957; Sperber & Wilson, 1986; Clark, 1996; Wilson & Wharton, 2006). For example, listeners use prosodic contrasts and situational context to distinguish ironic from sincere speech (Bryant & Fox Tree, 2002). As another example, listeners use intonation to make assessments about speakers' commitments to the content of their utterances (Brennan & Williams, 1995; Gunlogson, 2001; Smith & Clark, 1993). Steedman (2007) builds on Gunlogson's (2001) account, i.e. assessments of speaker commitment, by arguing that listeners use uptalk to make inferences about how speakers will proceed, actively assisting in the construction of higher level units of discourse. This account might better explain how uptalk (backward-looking and forward-looking), prolongations (delay), and contextual information (inferences about mental states) become integrated in on-line speech comprehension.

Many might dismiss the psycholinguistic examination of uptalk as excessive focus on a dialectal phenomenon. Like other elements of spontaneous speech, such as the choice of *like* as an enquoting device (Fox Tree & Tomlinson, 2008), uptalk might be an element of spontaneous speech that is noticeably in flux: more than a decade ago, researchers observed, "...recently we have heard undergraduates say things such as, 'So, y'know what? I have a professor, 'n he ...' where *professor* is spoken in the same intonation as it would be in the question, 'Have you ever seen my professor?'" (Gernsbacher & Jescheniak, 1995, p. 54). But regardless of whether uptalk is transient or not, it can be used currently to test pragmatic theories of intonation.

In conclusion, the interpretation of uptalk does not equal the sum of its parts. Similar to the difficulty of reconstructing the meaning of metaphors based on their parts, reconstructing the meaning of an intonational contour based on pitch accent and boundary tones can be misleading. Instead of a one-to-one form-function correspondence,

Table A1
Transcriptions for first utterance in pair – non-prolonged items.

Item transcription (boundary tone)	Pitch range (Hz)	F0 (Hz)	Duration (ms)
Non-prolonged rises			
This person was born on June 13th 1986 (L*H–H%)	298.29–168.62	238.83	308
Originally from Cedar Rapids, Iowa (H*L–H%)	307.82–148.74	277.99	240
I have a brother named Chris (H*L–H%)	296.28–161.58	224.06	372
He is five feet ten inches (L*H–H%)	414.28–248.74	241.39	257
I attended Immaculate Heart high school (H*L–H%)	249.77–140.43	201.73	272
My father was a college professor (H*L–H%)	380.87–140.47	290.14	301
I was born on June 13th, 1976 (L*H–H%)	228.88–80.31	220.36	342
I have three children (H*L–H%)	307.85–90.16	211.28	266
This person had two children (H*L–H%)	309.01–105.27	213.5	312
I was born on August 19th, 1946 (L*H–H%)	322.45–150.86	260.47	434
I was born on January, 17th 1962 (H*L–H%)	262.91–94.12	193.03	329
I am five foot seven (H*L–H%)	335.61–162.33	221.19	223
Non-prolonged falls			
She was born on November 4th, 1946 (H*H–L%)	349.64–200	266.62	255
I was born in Brooklyn, NY (H*L–L%)	299.34–175.73	201.99	363
I graduated from the Univ. of N. Carolina (H*L–L%)	339.95–153.72	208.35	242
He has a show on Comedy Central (H*H–L%)	311.25–186.16	224.68	227
His father was a teacher (H*L–L%)	346.67–199.47	224.4	286
My parents are divorced (H*L–L%)	139.87–87.17	98.13	356
I served in the US Army (H*H–L%)	306.1–2233	152.05	333
I have blond hair (H*L–L%)	222.46–149.34	188.27	379
I dropped out of college (H*L–L%)	286.08–159.92	182.35	291
She was born in Sherman Oaks (H*H–L%)	294.77–174.74	200.33	306
I was born on March, 19th, 1956 (H*H–L%)	316.1–204.4	217.67	303
I used to be a construction worker (H*L–L%)	330.14–175.7	213.97	269

Table A2
Transcriptions for first utterance in pair – prolonged items.

Item transcription (boundary tone)	Pitch range (Hz)	F0 (Hz)	Duration (ms)
Prolonged rises			
I was a Nickelodeon marketing executive (L*H–H%)	311.68–195.60	282.76	756
I am six foot two (H*L–H%)	289.87–121.97	245.58	564
I was born in Washington, DC (H*L–H%)	308.52–145.96	253.78	403
My birthday is February, 17th 1982 (H*L–H%)	287.35–108.05	217.63	533
He was born in Neptune, NJ (L*H–H%)	233.92–183.33	260.81	534
I was born January, 30th 1941 (L*H–H%)	262.87–129.54	197.79	369
I played a US Marine captain (H*L–H%)	296.35–174.54	237.39	415
My religious views are agnostic (L*H–H%)	323.39–142.87	287.99	541
I have a sister (L*H–H%)	254.68–154.13	234.13	573
I've been divorced twice (H*L–H%)	284.80–133.7	220.38	766
So, I drive a corvette (L*H–H%)	287.53–175.20	242.24	740
I've won an oscar (L*H–H%)	313.47–177.92	288.28	588
Prolonged falls			
She went to Tennessee State University (H*L–L%)	235.37–135.75	152.26	734
She's an actress (H*L–L%)	286.18–170.63	208.44	905
She was born in southern California (H*L–L%)	244.76–177.5	199.71	732
I was born on January 15th, 1929 (H*H–L%)	220.19–166.4	177.59	636
I was born on June 14th, 1986 (H*L–L%)	142.31–80.29	108.01	658
I'm a famous singer (L*H–L%)	126.87–80.0	96.12	449
I was born in Bay City, Michigan (H*L–L%)	128.29–74.06	88.78	477
I served in the Navy (H*H–L%)	260.89–147.2	157.01	587
I used to be a cocktail waitress (H*L–L%)	268.40–130.3	161.78	501
I have two sisters (H*H–L%)	280.21–190.2	208.25	495
She likes to play piano (H*L–L%)	282.62–179.1	183.7	565
I had a girlfriend named Kate Middelton (L*H–L%)	248.1–195.8	201.74	651

Note. F0 and duration are for the last syllable in the utterance. Manipulated versions of each token consisted of manipulating the direction of the boundary tone. For example H–L% went to H–H% or L–L% became L–H%, whereas L–H% went to L–L% or H–H% went to H–L%. A boundary tone (L% or H%) is defined by Pierrehumbert and Hirschberg (1990) as "the final tone in an intonational phrase" (p. 277), which is preceded by a phrasal accent (L- or H-).

the meaning of uptalk results from a complex interaction of time, presupposition, and inference. Given the complex nature of uptalk, it's no wonder everyone is talking it up.

Acknowledgments

This research was conducted for the first author's dissertation research at the University of California at Santa Cruz. The manuscript was written and submitted while he was a post-doctoral research associate at Stanford University and was supported by NSF Grant 0720054. The first author is now a post-doctoral research associate in the Psychology Department at Cardiff University. We thank the following research assistants who helped with this project: Lisa Jenson, Sarah Jerz, Ben Radosevich, Shane Hihara, Jeannine Mesina, and Jack Emery. We thank Jesse Snedeker, Natalia Blackwell, Kris Liu, and three anonymous reviewers for comments on the manuscript.

Appendix A

See Tables A1 and A2.

References

Allen, K. (1984). The high rise terminal contour. *Australian Journal of Linguistics, 4*, 19–32.

Anderson, A. H., Bader, M., Bard, E. H., Boyle, G. M., Doherty, S. C., Garrod, S., et al. (1991). The HRC map task corpus. *Language and Speech, 34*(4), 351–366.

Arnold, J. E., Hudson Kam, C. L., & Tanenhaus, M. K. (2007). If you say thee uh you are describing something hard: The on-line attribution of disfluency during reference comprehension. *Journal of Experimental Psychology: Learning, Memory and Cognition, 33*(5), 914–930.

Baayen, R. H. (2007). LanguageR: Data sets and functions with "analyzing linguistic data: A practical introduction to statistics" [computer software manual] (R package version 0.4).

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*(4), 390–412.

Bard, E. G., & Aylett, M. P. (1999). The dissociation of deaccenting, givenness and syntactic role in spontaneous speech. In J. Ohala, Y. Hasegawa, M. Ohala, D. Granville, & A. C. Bailey, (Eds.), *Proceedings of the XIVth international congress of phonetic science* (pp. 1753–1756). San Francisco.

Barr, D. J. (2001).Trouble in mind: Paralinguistic indices of effort and uncertainty in communication. In C. Cavé, I. Guaïtella, & S. Santi (Eds.) *Oralité et gestualité*: *Interactions et comportements multimodaux dans la communication* (pp. 597–600). Paris: L'Harmattan.

Barr, D. J. (2003). Paralinguistic correlates of conceptual structure. *Psychonomic Bulletin & Review, 10*, 462–467.

Bartels, C., & Kingston, J. (1994). Salient pitch cues in the perception of contrastive focus. In P. Bosch, R. van der Sandt, (Eds.), *Focus & natural language processing, proceedings of the journal of semantics conference on focus*, IBM working papers, TR-80.94-006.

Bates, D. (2007). lme4: Linear mixed-effects models using S4 classes [Computer software manual] (R package version 0.99875-9).

Baumann, S., & Grice, M. (2006). The intonation of accessibility. *Journal of Pragmatics, 38*(10), 1636–1675.

Beckman, M. E., & Ayers, G. M. (1994). Guidelines for ToBI labeling, version 3. Ohio State University.

Beckman, M. E., & Pierrehumbert, J. (1986). Intonational structure in English and Japanese. *Phonology Yearbook, 3*, 255–310.

Birch, S., & Clifton, C. Jr., (1995). Focus, accent, and argument structure: Effects on language comprehension. *Language and Speech, 33*, 365–391.

Bock, J. K., & Mazzella, J. R. (1983). Intonational marking of given and new information: Some consequences for comprehension. *Memory and Cognition, 11*, 64–76.

Boersma, P., & Weenink, D. (2008). Praat: Doing phonetics by computer (version 5.0.08) [computer program]. <http://www.praat.org/> Retrieved 11.02.08.

Brennan, S. E., & Williams, M. (1995). The feeling of another's knowing: Prosody and filled pauses as cues to listeners about the metacognitive states of speakers. *Journal of Memory and Language, 34*(3), 383–398.

Bryant, G. A., & Fox Tree, J. E. (2002). Recognizing verbal irony in spontaneous speech. *Metaphor and Symbol, 17*(2), 99–117.

Byrd, D., & Saltzman, E. (2003). The elastic phrase: Modeling the dynamics of boundary-adjacent lengthening. *Journal of Phonetics, 31*, 149–180.

Chafe, W. L. (1976). Givenness, contrastiveness, definiteness, subjects, topics and point of view. In C. N. Li (Ed.), *Subject and topic* (pp. 27–55). New York: Academic Press.

Clark, H. H. (1996). *Using language*. Cambridge, UK: Cambridge University Press.

Clark, H. H. (2002). Speaking in time. *Speech Communication, 36*, 5–13.

Clark, H. H., & Bangerter, A. (2004). Changing conceptions of reference. In I. Noveck & D. Sperber (Eds.), *Experimental pragmatics* (pp. 25–49). Basingstoke, England: Palgrave Macmillan.

Clark, H. H., & Fox Tree, J. E. (2002). Using uh and um in spontaneous speech. *Cognition, 84*, 73–111.

Clark, H. H., & Haviland, S. E. (1977). Comprehension and the given-new contract. In R. O. Freedle (Ed.), *Discourse production and comprehension* (pp. 1–40). Norwood, NJ: Ablex.

Cruttenden, A. (1997). *Intonation* (2nd ed.). Cambridge: Cambridge University Press.

Cutler, A. (1976). Phoneme-monitoring reaction time as a function of preceding intonation contour. *Perception & Psychophysics, 20*(1), 55–60.

Cutler, A., & Clifton, C. Jr., (1984). The use of prosodic information in word recognition. In H. Bouma & D. G. Bouwhnis (Eds.), *Attention and performance X* (pp. 183–196). Hillsdale, NJ: Erlbaum.

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

Dunn, A. L., & Fox Tree, J. E. (2009). A quick, gradient bilingual dominance scale. *Bilingualism: Language and Cognition, 12*(3), 273–289.

Ferreira, F. (1993). Creation of prosody during sentence production. *Psychological Review, 100*, 233–253.

Fletcher, J., Stirling, L., Mushin, I., & Wales, R. (2002). Intonational rises and dialog acts in the Australian English map task. *Language and Speech, 45*(3), 229–252.

Fox Tree, J. E. (2001). Listeners' uses of um and uh in speech comprehension. *Memory and Cognition, 29*(2), 320–326.

Fox Tree, J. E. (2010). Discourse markers across speakers and settings. *Language and Linguistics Compass, 3*(1), 1–13.

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

Fox Tree, J. E., & Tomlinson, J. M. (2008). The rise of like in spontaneous quotations. *Discourse Processes, 45*(1), 85–102.

Garrod, S., & Sanford, A. J. (1982). Bridging inferences in the extended domain of reference. In A. Baddeley & J. Long (Eds.), *Attention and performance IX* (pp. 331–346). Hillsdale, N.J.: LEA.

Gee, P. G., & Grosjean, F. (1983). Performance structures: A psycholinguistic and linguistic appraisal. *Cognitive Psychology, 15*, 411–458.

Gernsbacher, M. A., & Jescheniak, J. D. (1995). Cataphoric devices in spoken discourse. *Cognitive Psychology, 29*(1), 24–58.

Gravano, A., Benus, S., Hirschberg, J., Sneed, E., & Ward, G. (2008). The effect of contour type and epistemic modality on the assessment of speaker certainty. Paper presented at the proc. international conference: Speech prosody 2008, Campinas, Brazil.

Grice, P. (1957). Meaning. *Philosophical Review, 66*, 377–388.

Guhe, M., Steedman, M., Bard, E. G., & Louwerse, M. M. (2006). Prosodic marking of contrasts in information structure. In *Proceedings of BranDial 2006*: *The 10th workshop on the semantics and pragmatics of dialogue*, University of Potsdam, Germany, September 11th–13th, 2006.

Gunlogson, C. A. (2001). *True to form*: *Rising and falling declaratives as questions in English*. Unpublished Doctoral Dissertation, University of California Santa Cruz.

Hirschberg, J., & Pierrehumbert, J. (1986). The intonational structuring of discourse. In *Proceedings of the 24th annual meeting on association for computational linguistics*. (pp. 136–144). New York.

Hirschberg, J., & Ward, G. (1992). The influence of pitch range, duration, amplitude, and spectral features on the interpretation of the rise-fall-rise contour in English. *Journal of Phonetics, 20*, 241–251.

House, J. (2007). The role of prosody in constraining context selection: A procedural approach. *Nouveux Cahiers de Linguistique Française, 28*, 369–383.

Ito, K., & Speer, S. R. (2008). Anticipatory effects of intonation: Eye movements during instructed visual search. *Journal of Memory and Language, 58*(2), 541–573.

Kohler, K. J. (2004). *Prosody revisited*: *Function, time, and the listener in intonational phonology*. Paper presented at the proc. international conference: Speech prosody 2004, Nara, Japan.

Kohler, K. (2006). Paradigms in experimental prosodic analysis: From measurement to function. In S. Sudhoff, D. Lenertovà, R. Meyer, S. Pappert, P. Augurzky, & I. Mleinek, et al. (Eds.), *Methods in empirical prosody research* (pp. 123–152). Berlin: Walter de Gruyter.

Kowtko, J. (1996). *The function of intonation in task-oriented dialogues*. Ph.D thesis, University of Edinburgh.

Krahmer, E., & Swerts, M. (2001). On the alleged existence of contrastive accents. *Speech Communication, 34*(4), 391–405.

Krifka, M. (2007). Basic notions of information structure. In C. Féry, G. Fanselow, & M. Krifka (Eds.), *The notions of information structure* (pp. 13–55).

Ladd, D. R., & Morton, R. (1997). The perception of intonational emphasis: Continuous or categorical? *Journal of Phonetics, 25*(3), 313–342.

Liberman, M. (2006). Uptalk is not HRT. Language log. <http://itre.cis.upenn.edu/~myl/languagelog/archives/002967.html> Retrieved 30.03.06.

McLemore, C. A. (1991). *The pragmatic interpretation of English intonation*: *Sonority speech*. Unpublished Doctoral Dissertation, University of Texas at Austin.

Ohala, J., Dunn, A., & Sprouse, R. (2004). *Prosody and Phonology*. Paper presented at the Speech Prosody, Nara, Japan.

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences, 27*, 169–225.

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

Prince, E. F. (1981). Toward a taxonomy of given-new information. In P. Cole (Ed.), *Radical pragmatics* (pp. 223–256). New York: Academic Press.

Quené, H., & van den Bergh, H. (2008). Examples of mixed-effects modeling with crossed random effects and with binomial data. *Journal of Memory and Language, 59*(4), 413–425.

Ramus, F., & Mehler, J. (1999). Language identification with suprasegmental cues: A study based on speech synthesis. *Journal of the Acoustic Society of America, 105*(1), 512–521.

R Development Core Team. (2007). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. <http://www.R-project.org/> (ISBN 3-900051-07-0).

Richardson, D. C., Dale, R., & Tomlinson, J. M. Jr., (2009). Conversation, gaze coordination and beliefs about visual context. *Cognitive Science, 33*, 1468–1482.

Selkirk, E. (1995). Sentence prosody: Intonation, stress, and phrasing. In J. Goldsmith (Ed.), *The handbook of phonological theory* (pp. 550–569). Cambridge, MA: Blackwell.

Shokier, V. (2008). Evidence for the stable use of uptalk in South Ontraio English. *Working Papers in UPenn Linguistics, 14*(2), 16–21.

Smith, V. L., & Clark, H. H. (1993). On the course of answering questions. *Journal of Memory and Language, 32*(1), 25–38.

Sperber, D., & Wilson, D. (1986). *Relevance: Communication and cognition*. Cambridge, MA: Blackwell, Oxford and Harvard University Press.

Steedman, M. (2007). Information-structural semantics for English intonation. In C. Lee et al. (Eds.), *Topic and focus: Cross-linguistic perspectives on meaning and intonation* (pp. 254–264). Dordrecht: Springer Verlag.

Terken, J., & Nooteboom, S. G. (1987). Opposite effects of accentuation and deaccentuation on verification latencies for given and new information. *Language and Cognitive Processes, 2*, 145–163.

Ward, N., & Tsukahara, W. (2000). Prosodic features which cue back-channel responses in English and Japanese. *Journal of Pragmatics, 32*, 1177–1207.

Watson, D., & Gibson, E. (2004). The relationship between intonational phrasing and syntactic structure in language production. *Language and Cognitive Processes, 19*(6), 713–755.

Watson, D., Gunlogson, C. A., & Tanenhaus, M. K. (2006). Online methods for investigation of prosody. In S. Sudhoff, D. Lenertova, R. Meyer, S. Pappert, P. Augurzky, I. Mleinek, N. Richter, & J. Schlieber (Eds.), *Methods in empirical prosody research* (pp. 259–282). New York: Walter de Gruyter.

Watson, D. G., Tanenhaus, M. K., & Gunlogson, C. A. (2008). Interpreting pitch accents in online comprehension: H^* vs. $L + H^*$. *Cognitive Science, 32*(7), 1232–1244.

Wilson, D., & Wharton, T. (2006). Relevance and prosody. *Journal of Pragmatics, 38*(10), 1559–1579.