**Processing intonationally implicated contrast versus negation in American English**

Heeyeon Dennison
Defense Language Institute Foreign Language Center

Amy J. Schafer
University of Hawaii

Corresponding author:
Amy J. Schafer
Department of Linguistics
University of Hawaii
1890 East-West Rd
Honolulu, HI 96822
Email: amy.schafer@hawaii.edu

**Abstract**

Certain English intonational contours facilitate a conversational implicature that a relevant alternative to the stated proposition does not hold true. We evaluated how frequently and how quickly naïve participants achieved such pragmatically enriched meanings when their attention had not already been drawn to a set of alternatives. Sentences with L+H* L-H% intonational contours, along with broad focus affirmative and negative counterparts, were tested in a pair of experiments. Experiment 1 revealed that most interpretations of the L+H* L-H% sentences evidenced the expected implicature, but a substantial number did not. Experiment 2 mapped the activation levels across time for the asserted state and a contradictory/implicated alternative for the same three sentence types, using a picture-naming paradigm. The results revealed that lexical negation produced a contrast in activation levels between the two alternatives at an earlier time point than the L+H* L-H% contour, and that the relative activation of the two states shifted over time for L+H* L-H% sentences, such that an intonationally implicated alternative was highly activated at a time point when the activation for the asserted meaning had declined. These results further our understanding of the pragmatic processes involved in the interpretation of negation and intonation.

## Introduction

It has long been observed that intonation can encourage the listener to draw an interpretation that goes beyond the literal expression. For example, a speaker who has been asked about his exam performance can use the right pronunciation of *I passed* to highlight that he merely passed and did not perform at some better level (Rooth, 1992). In particular, tunes with rising pitch accents coupled with a phrase-final rise are known for their ability to convey that the listener should generate the implicature that a contextually plausible alternative is not true, although researchers have differed on how to best characterize the form-meaning relationships (Bing, 1979; Bolinger, 1958; Büring, 2007; Constant, 2012; Hirschberg & Ward, 1992; Jackendoff, 1972; Ladd, 2008; Lee, 2000; Pierrehumbert & Hirschberg, 1990; Steedman 2000; Wagner, 2012; Ward & Hirschberg, 1988, among others).

Relatively little work has investigated how such tunes are comprehended by naïve listeners, particularly when a set of alternatives have not already been made salient by a preceding visual or linguistic context. In the present study, we investigated the processing of such intonationally rich sentences when presented in isolation, to better understand the specific contribution of the intonational form in the comprehension of the target meaning. It is widely understood that intonation is used to connect propositional content to the information state that has been constructed by the speaker and listener in a particular discourse situation (Pierrehumbert & Hirschberg, 1990). This suggests that intonational meaning is highly dependent on contextual information. However, certain intonational forms are also frequently discussed as being associated with specific functions (e.g., introducing a new topic, evoking contrast, expressing contradiction), which can foster a (misleading) view that particular tunes convey a single meaning, and can misrepresent the processing space for spoken sentences.

Therefore, we were first interested in documenting how frequently listeners adopt different interpretations for sentences pronounced with a specific intonational form (i.e., our target tune, as explained below). In addition, we examined the apparent time course for the activation of a basic versus enriched meaning in response to the intonational form, for sentences presented without the support or hindrance of additional contextual information. To situate the intonational findings in the processing literature, we compared the processing of sentences that implicate the negation of an alternative through their intonational form to ones with explicit negation. The study thus provides new findings on two types of sentences that highlight the contrast between alternative propositions, and to our knowledge, the first direct comparison of the online processing of lexical negation versus intonationally implicated contrast.

Specifically, we examined sentences such as (1), realized with an unaccented subject, a rising accent on *was*, and a rising phrase-final contour. Accentuation on *was* supports a contrast between the asserted state and an alternative state, and so for convenience we will refer to sentences with our target tune as Contrastive sentences. Our expectation was that these sentences would lead to more than one type of interpretation. The interpretation of primary interest involves an implicature: for (1), that the mailbox is no longer full. Under this interpretation the asserted state is set against alternatives (e.g., *was full then, is full now)*. A listener can infer from the indication of contrast that alternatives (e.g., *was full then and is still full now*) do not hold, especially if an alternative such as *is full* is more likely to convey a relevant proposition in the discourse environment (Grice, 1975). We will call this an implicature of state contrast. The target implicature for (1) is likely facilitated in a discourse context in which relevant alternatives to the assertion (e.g., a now-empty mailbox) are presupposed (Roberts, 1996/2012). Therefore, we presented sentences like (1) without a preceding visual or linguistic context that established

alternatives, to provide a rigorous test of the contribution of the intonational form to the ultimate interpretation.

(1) The mailbox WAS full…

Numerous psycholinguistic experiments have revealed rapid and significant effects of intonational information in discourse processing. These include studies showing effects of phrase-final tones (e.g., Heeren, Bibyk, Gunlogson, & Tanenhaus, 2015; Kjelgaard & Speer, 1999) and of pitch accents that evoke a contrastive interpretation (e.g., Dahan, Tanenhaus, & Chambers, 2002; Ito & Speer, 2008). Nevertheless, the research on accentuation and information structure has drawn heavily on situations in which alternatives were established prior to the presentation of the critical spoken material, such as experiments in which a visual display of blue versus green objects preceded the presentation of sentences with accented color adjectives. This leaves open the question of how particular tunes will affect processing when the listener must generate a plausible set of alternatives in his mind – as happens in many natural discourse situations.

Research on inferential processing, such as the interpretation of *some* as 'some but not all' and not merely 'at least one' has also raised questions about how readily different types of inferences are drawn across different discourse or testing situations. This research has shown that enriched interpretations can require more processing time than computations of the basic semantic meaning (e.g., Bott & Noveck, 2004; Bott, Bailey, & Grodner, 2012; Huang & Snedeker, 2009, 2011), especially when the inference has not been boosted by contextual support (Degen & Tanenhaus, 2015) or conventionalized use (Levinson, 2000).

Prior work on intonation and implicature has not yet made clear how strong the associations might be between particular tunes and particular types of implicatures, although this work has begun to demonstrate intonational effects on enriched interpretations. Chevallier, Noveck, Nazir, Bott, Lanzetti & Sperber (2008) established that the use of a pitch accent on the French disjunctive connector encouraged French listeners to compute an enriched, exclusive-*or* interpretation ('but not both') in a verification task, and argued that the steps involved in drawing this inference can require measurable processing time. Research on American English sentences more similar to (1), using the visual world paradigm, has demonstrated that the intonational contour supports an implicature to negate an alternative, yet suggests a need for further consideration of how various factors contribute and interact, a question we return to below (Dennison, 2010; Dennison & Schafer, 2010; Kurumada, Brown, Bibyk, Ponillo & Tanenhaus, 2014a). Importantly, these visual world studies as well as a set of related offline studies by Kurumada, Brown, and Tanenhaus (2012) and the other judgment studies have all involved some kind of context that established a set of alternatives, and so do not resolve how strongly the intonation form itself contributes to the target interpretation.

We made use of a pair of tasks that began each trial with the presentation of a null-context spoken stimulus. Experiment 1 collected open-ended continuations and naturalness judgments. Experiment 2 employed a picture-naming paradigm. In this latter task we first presented the spoken sentence, and then collected the naming time for an image consistent with the basic, asserted meaning (a full mailbox, for (1)) or for an implicated alternative (an empty mailbox). By varying the inter-stimulus interval (ISI) between the offset of the sentence and the onset of picture, we were able to estimate the relative level of activation of each state across time. This allowed us to evaluate how quickly listeners derive the pragmatically enriched

meaning in the absence of a rich context, i.e., on the basis of the lexical and intonational information alone.

Both tasks tested Contrastive sentences like (1) along with two other sentence types: broad focus affirmative sentences and broad focus negated sentences (e.g., *The mailbox was (not) full)*. The broad focus affirmative sentences provided a sensible baseline for comparison since they matched the Contrastive sentences in every aspect except the intonational form. The broad focus negative sentences were included because the targeted enriched meaning of the Contrastive sentences implicates a change from the asserted state to the implicated state and so, we assumed, would involve activation of two different alternatives from the continuum established by the predicate adjective (e.g., 'full', 'not full'). Negated sentences such as *The mailbox is not full* similarly involve a contrast between two states. They are commonly used in discourse situations in which the counterfactual state (a full mailbox) is presupposed or questioned, and the negation of it (a non-full mailbox) is being asserted. Several previous studies have investigated the time course of activation for factual versus counterfactual states with negated sentences, providing a foundation for comparison with the Contrastive sentences (e.g., Anderson, Huette, Matlock, & Spivey, 2010; Kaup, Yaxley, Madden, Zwaan, & Lüdtke, 2007; Tian, Ferguson, & Breheny, 2016). The negated sentences are especially relevant in Experiment 2, so we postpone further review of the processing of negation until the introduction to that experiment.

Research on intonation and implicated meaning has been complicated by the difficulty of describing the intonation associated with the critical meaning and by heterogeneous traditions in terminology across sub-fields and theoretical traditions. Early research on intonation and implicature used a range of descriptions for tunes that support our target meaning. Using the

annotations of the ToBI system (Beckman & Ayers, 1997; Veilleux, Shattuck-Hufnagel, & Brugos, 2006), researchers have argued that the L*+H L-H% tune, also called the rise-fall-rise contour, readily evokes enriched meanings, yet it is also known that subtle changes in it can shift the interpretation (Constant, 2012; Hirschberg & Ward, 1992; Ward & Hirschberg 1985, 1988). Other work (Lee, 2000; Kurumada et al., 2014a; Roberts, 1996/2012) supports the claim that our target implicature also arises with an L+H* accent on the verb and a final L-H%, which is the transcription we give to our stimuli. The L+H* L-H% tune is commonly described as marking a contrastive topic (e.g., Büring, 2007; Lee, 2000) or theme (Steedman, 2000); see Constant (2012) for discussion of semantic connections among these. Roberts (1996/2012) generalizes across multiple uses of the tune by describing the pitch accent as contributing to the salience of alternatives and the final rise as indicating that there is more that could be said, echoing a compositional view presented by Pierrehumbert & Hirschberg (1990).

The distinctions among (potential) categories of English pitch accents have not been completely clear-cut (e.g., Ladd, 1983; Ladd & Schepman, 2003; Pierrehumbert & Steele, 1989), and appear to differ across dialects (Arvaniti & Garding, 2007). There may well be an acoustic continuum between L+H* rises and L*+H rises and variability in the interpretation of specific tokens. Pierrehumbert & Hirschberg (1990) argue that both rising accents evoke a scale of alternatives. The broader intonational literature notes that we often see multiple intonational realizations used to convey a single meaning and that multiple meanings are associated to a single intonational form – not because speakers are unreliable in their production or listeners err in their comprehension, but because of the complexity of intonation-to-meaning pairings (e.g., Warren, 2016). We do not want to claim that the contour we test is the only tune that supports enriched interpretations. Rather, we aimed to further characterize the role of intonation of

American English by providing empirical evidence about the meanings and activation patterns generated in response to sentences with one relevant tune, the L+H* L-H% contour. We now turn to Experiment 1.

**Experiment 1**

Experiment 1 examined the continuations generated for sentences like *The pencil WAS sharp*, pronounced with an L+H* L-H% tune, and our two types of baseline conditions: affirmative and negative sentences pronounced with a broad focus intonational pattern (see Table 1 and Figure 1). Participants listened to sentences spoken in one of these forms and created a follow-up sentence or clause they thought would naturally continue the discourse, using a common paradigm in experimental studies of discourse (e.g., Kehler, Kertz, Rohde & Elman, 2008; Arnold, 2001). We reasoned that if participants derived an implicature of state contrast they would produce continuations consistent with that interpretation, so that the proportion of continuations expressing or presupposing state contrast would be a reasonable estimate of how frequently participants reached the targeted enriched interpretation of the spoken sentences we presented. Unlike forced-choice tasks in which participants are given interpretations and explicitly asked to choose their preferred response, our open-ended continuation task with null contexts did not make any particular interpretations salient to the participant. Therefore, it was well-suited to our primary research question for Experiment 1, of the extent to which the sentential information alone leads to the pragmatically enriched interpretation.

INSERT TABLE 1 ABOUT HERE

INSERT FIGURE 1 ABOUT HERE

We predicted that the L+H* L-H% tune of the Contrastive condition would lead to significantly more cases of the targeted implicature than the other two conditions, as might be expected from earlier research on this contour. Thus one goal was to verify that we would find frequent generation of state-contrast implicatures in the Contrastive condition, despite the null context. However, we also anticipated that listeners would not always generate a state-contrast implicature in this condition. First, the intonational literature, and our own intuitions, suggested alternative interpretations (Constant, 2012; Steedman, 2000). Recall that in common approaches to intonational meaning the L+H* accent invites contrast with alternatives from a situationally relevant set, and the final rise indicates some type of incompleteness (Pierrehumbert & Hirschberg, 1990; see also Warren, 2016). Such an approach is compatible with the tune being used with our stimuli when the speaker wishes to express verum focus (to emphasize that the pencil was, indeed, sharp) and has some other reason to express non-finality, such as if the speaker plans to immediately continue with closely connected information. An example of this (from the data collected in Experiment 1) is: *The belt WAS buckled, so my pants stayed on*. Second, it might be difficult for listeners to adopt an enriched interpretation when there is no preceding context that supplies a relevant set of alternatives for it. Therefore, we predicted that not all continuations of the Contrastive condition would reflect an implicated state contrast.

The H* H* L-L% tune used for the Affirmative sentences is often used to introduce new, broadly focused information into the discourse, and so we anticipated that listeners would focus more on the asserted state of the entity, rather than any contrastive state, and would thus produce few continuations expressing state contrast. However, because of the ready contrast that binary

predicates allow, we predicted that even Affirmative sentences might generate some continuations indicating state contrast, but that the probability of such continuations would be much lower than in the Contrastive condition. Negation can also evoke contrastive states, since it is used to indicate the discrepancy between a presupposed expectation and a factual state. Nevertheless, we predicted that our broad-focus negative sentences would encourage participants to focus on what was asserted (e.g., that the pencil was not sharp) and to continue with reasons why the presupposition (that the pencil was sharp) was not met (e.g., *because I used it all day*) or with consequences of the negated state (e.g., *so I had to find a pen*) instead of changes in state. We thus expected that the proportion of continuations indicating state-contrast implicature would be lower after Negative sentences than Contrastive ones. The key question was not whether the conditions would differ from each other but by how much, and how frequently the state-contrast meaning would be generated in the Contrastive condition.

*Participants*

Eighteen native English speakers from the University of Hawai‘i community participated either for course credit or $5 compensation.

*Materials and design*

Experiment 1 tested 33 critical sentences, each recorded in all three conditions shown in Table 1 above: Contrastive, Affirmative, and Negative. All target sentences included a definite subject noun phrase, the past auxiliary *was*, and a predicate adjective. The predicate adjectives described binary attributes, with meanings that were contradictory (e.g., *sharp* vs. *dull*), contrary (*full* vs. *empty*), or reversible (*tied* vs. *untied*) and represent the ends of a scale or pole (Gross,

Fischer & Miller, 1989; Israel, 2004). These were chosen to facilitate processing and coding of state-contrast interpretations, because accessing one end of the scale of a set of binary attributes allows easy access to the opposite end (Deese, 1964; Frazier, Clifton, & Stolterfoht, 2008; Gross et al., 1989; Kaup, Lüdtke, & Zwaan, 2006; Kennedy & McNally, 2005). See Appendix A for the list of critical sentences.

All sentences were recorded by a female phonetician proficient in the Mainstream American English ToBI system. Acoustic measurements of the duration and pitch excursions (Tables 2 and 3 below, respectively), along with ToBI annotations by two independent transcribers, verified that the sentences were produced as intended. In particular, the F0 of the Contrastive condition fell gradually during the subject phrase to a low point realized at the end of the subject phrase. The F0 rose during the accented auxiliary *WAS*, peaking prior to its end, establishing an unambiguously rising accent. F0 was then lower on the subsequent word, which ended with a clear sentence-final rise. Inferential statistics from the acoustical analyses of the stimuli appear in Appendix B. Sample soundfiles are available at http://www2.hawaii.edu/~aschafer/.

INSERT TABLES 2 AND 3 ABOUT HERE

In addition to the 33 test sentences, 20 filler sentences were prepared with the same syntactic forms as the critical items (16 affirmative and 4 negative sentences), but variation in intonational form and predicate type. Alternative intonational contours (e.g., L* H-H%, where L* was placed on the subject phrase) were included because we were interested in the interpretations generated when the critical intonation was mixed with a range of other tunes, as

happens in natural settings. A more diverse set of predicate adjectives (e.g., *minty*, *purple*, *prickly*) was employed in fillers to neutralize any potential expectations for contrast or binary scales from the lexical properties of predicates. In addition, 12 of the 20 fillers intentionally contained minor mistakes (e.g., unwarranted pauses, mispronunciation) because the procedure also asked participants to rate the naturalness of the stimuli and we wanted that to be a meaningful test. The 33 critical sentences and 20 fillers were distributed across three running lists in a Latin Square design, so that each participant received only one condition for each critical item and was tested on all 53 items.

*Procedure*

Participants listened to one sentence at a time over headphones, and rated its naturalness on a five-point Likert scale. The sentence was then replayed, after which the participant could change the naturalness score if desired (a rare occurrence), and typed a continuation. Participants were told to do this as quickly as possible, providing whatever continuation came first to mind. Each experiment began with four practice trials and used a different random presentation order for each participant. The entire experiment session took approximately 20 minutes.

*Results*

Analyses of the naturalness judgments indicated that all critical stimuli were sufficiently natural. Of the 594 critical trials, there were four for which participants omitted continuations; these were removed from subsequent analysis. Three native English speakers independently coded each continuation into one of four categories described below. Fourteen trials (less than 3% of the total data) were eliminated for not generating agreement from at least two of the three

coders; all three coders agreed for 435 out of the 590 trials that received a continuation (74%). The coders classified the continuation meaning into one of four distinct categories: (a) state contrast, (b) other contrast, (b) neutral, and (d) other. Given a target sentence like *The pencil was sharp*, the coder utilized the *state contrast* category when the continuation expressed a meaning that contradicted the state of affairs asserted in the target sentence and included an alternative state (e.g., …*but now it's dull)*. However, if the continuation expressed contrast with respect to other parts of the sentence such as the subject (e.g., …*but the pen was useless*) or the entire proposition (…*but I still couldn't poke a hole in the paper cup*), the coders marked it as *other contrast*. A continuation was coded as *neutral* when it expressed a meaning that accepted the state of affairs asserted in the test sentence and continued it without any obvious or salient contrast (e.g., …*and it was useful for filling out bubbles on the form*). The last category, *other*, included any continuations that coders thought of as ambiguous or non sequiturs (e.g., *The door was ajar.* → *Of course the door isn't a jar.*).

Over 90% of the data were coded in either the state contrast or neutral category, and less than 2% of the data were coded as expressing non-state contrast. Therefore, the data were analyzed using mixed-effects logistic regression on state contrast versus all other types of interpretations, with the maximal random effects structure permitted by the data. The Contrastive condition induced continuations expressing state contrast significantly more often (63.3% of the trials) than both the Affirmative and Negative conditions (7.2% and 1.5% of the trials, respectively; $\beta = -4.07$, $z = -9.3$, $p < .05$; $\beta = -5.78$, $z = -8.4$, $p < .05$). That is, the Contrastive condition—but not the other conditions—frequently resulted in continuations consistent with the implicated state contrast despite the absence of a preceding facilitative context.

An important finding was that the Contrastive condition resulted in state-contrast continuations in notably less than 100% of the trials. This condition generated neutral meaning for 23.4% of the trials, and all other types of contrastive meaning (besides state contrast) were found in only 2.7% of the trials. This result reinforces the existing observation that one type of intonational tune can serve multiple functions (e.g., Gussenhoven, 2004; Ladd, 2008; Pierrehumbert & Hirschberg, 1990).

*Discussion*

Experiment 1 established empirical evidence from naïve participants that the L+H* L-H% tune can frequently invite continuations consistent with a state-contrast implicature, and revealed that such continuations were the most common continuation for our specific test situation.[1] This was the case even though the sentences were presented with no preceding linguistic or visual context to support the enriched interpretation, using a task that simply asked participants to generate continuations, i.e., a task that did nothing to highlight particular alternatives. The intonational pattern therefore appears to provide strong support for the target implicature. The Contrastive condition evoked state-contrast interpretations substantially more often than the Affirmative and Negative sentences. These results demonstrate how important intonation can be to sentence interpretation, as dramatically different patterns of continuations were generated for the Contrastive condition and the Affirmative condition, which were identical except for the intonational form. Nevertheless, the target implicature was generated in only 63% of the Contrastive trials. It was far from an automatic consequence of the tune. This finding is consistent with a one-to-many mapping from an intonated sentence to interpretations of it. Such variability in interpretation can lead to situations in which a speaker's intended meaning is not

(or at least not initially) constructed by the comprehender, underlining the importance of investigating intonational meaning across a range of discourse situations (including situations resulting in miscommunication). Experiment 2 extended our investigation to the online processing of these three sentence types and the L+H* L-H% tune.

**Experiment 2**

Experiment 2 was designed to evaluate how quickly listeners derive a state-contrast implicature from L+H* L-H% sentences like (1), by assessing activation for the asserted versus implicated state at multiple time points. As in Experiment 1, we were interested in the independent contribution of the intonational information on meaning computation when alternative propositions had not yet been established or presupposed in the discourse situation. Thus, we presented our target sentences without any preceding context and compared them to broad-focus affirmative and negative counterparts. Previous research investigating the time course of processing negative versus affirmative sentences has argued that it can take longer for a null-context negative assertion to be processed than an affirmative one, because of additional processing steps involved in the interpretation of negated sentences (e.g., Kaup et al., 2006; Tian et al., 2016). However, the exact processing mechanisms and representations for negation have been the subject of debate, and the processing of negation is known to differ across discourse contexts (e.g., Dale & Duran, 2011; Giora, 2006; Huette, 2016; Kaup, et al., 2007; Nieuwland & Kuperberg, 2008; Orenes, Beltrán, & Santamaría, 2014; Tian et al., 2016). Given that the processing of Contrastive sentences like (1) involves negating alternatives (at least implicitly), we first summarize existing research on activating alternative states with negative versus

affirmative sentences. Then, we review two visual-world studies that also tested our target contour.

*Processing negated sentences*

Previous research using picture-naming and probe-recognition tasks has demonstrated that images consistent with the interpretation of just-received linguistic material result in shorter reaction times than inconsistent images (e.g., Kaup et al., 2006, 2007; Zwaan, Stanfield & Yaxley, 2002). Notably, Kaup et al. (2006) used a picture-naming task with the German equivalents of affirmative and negative binary-predicate sentences such as *The door is (not) open* to investigate the time point at which a match-facilitation effect would emerge in each sentence type. In their study, participants read the critical sentence and then saw an image of an open or a closed door, at 750-ms or 1500-ms ISIs. Participants then named the depicted object (e.g., "door") as quickly as possible. Kaup et al. found that after affirmative sentences, picture-naming times at the 750-ms ISI were significantly shorter for images depicting the mentioned state than the opposite state. However, after negative sentences, significant match-facilitation for the factual state (shorter naming times for a closed door after processing *The door is not open*) did not emerge until the 1500-ms ISI. Because Image type was not significant for affirmative sentences at the 1500-ms ISI or negative ones at the 750-ms ISI, Kaup et al. claimed that their results demonstrate a match-facilitation effect for the final interpretation of a sentence and suggested that comprehenders reach the final interpretation more quickly with affirmative sentences than negative ones.

Anderson et al. (2010) conducted a similar experiment using written English affirmative and negative sentences with binary predicates and a 1000-ms ISI, and found match effects for

both affirmative sentences (an advantage for the mentioned state) and negative sentences (an advantage for the factual over the counterfactual state). Note, however, that Anderson et al. employed a probe recognition task in which participants evaluated whether the depicted object had been mentioned in the previous sentence or not. Such a task involves a metalinguistic component that might induce participants to engage in greater use of mental imagery than they would employ during ordinary language use, in anticipation of the visual probe and the need for a decision about it. Probe recognition tasks involving negative sentences have nevertheless been important to the debates about how negation is processed, as these tasks have shown that at short ISIs, certain negative sentences result in significant decision-time advantages for images that match the counterfactual state: a mismatch advantage (e.g., Kaup et al., 2007; Tian et al., 2010). Initial work from this paradigm was taken as support for a two-stage model of negation, in which comprehenders first process the non-negated proposition and then turn to the negated interpretation (e.g., Kaup et al., 2007).

Many other studies have found greater difficulty for processing negative sentences than affirmative sentences (e.g., Carpenter & Just, 1975; Clark & Chase, 1972; Trabasso, Rollins, & Shaughnessy, 1971). Importantly, though, negative sentences have been shown to be no more difficult than affirmative sentences when given appropriate pragmatic support (e.g., De Villiers & Flusberg, 1975; Glenberg, Robertson, Jansen, & Johnson-Glenberg, 1999; Lüdtke & Kaup, 2006; Nieuwland & Kuperberg, 2008). And, the match versus mismatch effects have varied with the exact type of negative sentence (Tian et al., 2010; Tian, 2014), a point we return to below.

Recent work using the visual-world paradigm has also demonstrated contextual effects on the processing of negation, in addition to providing further evidence of longer processing times for negative versus affirmative sentences in certain contexts. Orenes et al. (2014) tested spoken

Spanish equivalents of sentences like *The figure is (not) red* in contexts which either restricted the set of alternatives to two colors or allowed multiple options. Gaze patterns differed by context. In the binary context, the proportion of looks to an image consistent with the mentioned state in affirmative sentences became significantly different from looks to the relevant alternative just 380 ms after the onset of the critical color word, whereas the differentiation point for the negative sentences did not occur until 1340 ms after the critical-word onset. In the multiple-option context, both the affirmative and the negative sentences produced increases in looks to the mentioned color at short lags from the onset of the color word, which indicated that the negative sentences resulted in a preference for fixations on the image that matched the counterfactual state (a mismatch advantage). These mismatch fixations from negative sentences increased for approximately 250 ms and then declined. This is similar to a mismatch advantage seen at early ISIs in probe-recognition studies with some types of negative stimuli – namely, stimuli in which the negative sentence did not, in the null discourse contexts of the experiments, easily support a specific alternative (Kaup et al., 2007; Tian et al., 2010; Tian, 2014).[2] Orenes et al. proposed that in a binary context comprehenders might maintain a representation of the factual state, while in the multiple-option context it might be easier to maintain a representation that contains negation, perhaps via a symbolic mental tag (e.g., Carpenter & Just, 1975; Johnson-Laird, 2001; see Orenes et al. and Tian et al. (2016) for further discussion of proposed models for the mental representation of negation.)

Tian et al. (2016) conducted a visual-world study that manipulated the syntactic form of affirmative and negative sentences, comparing simple forms to clefted ones like *It is John who has(n't) ironed his brother's shirt*. They assumed that the simple versus clefted negatives involve different presuppositions, such that the simple negatives presuppose the question of whether the

event has happened or not, but the clefted negatives address the question of who it is that hasn't performed the event. Tian et al. (2010, 2016) argued that listeners process sentences with respect to such Questions Under Discussion (Roberts, 1996/2012) and so processing the simple negative involves accessing a representation of the counterfactual alternative (that John has ironed his brother's shirt) whereas processing the negative cleft does not. They predicted that these different patterns of pragmatically-based meaning activation would lead to distinct patterns of eye fixations for the simple sentences versus the clefts, which is indeed what they found, as described below.[3]

Importantly, Tian et al. (2016) created a discourse context expected to facilitate activation of the factual versus counterfactual states for the simple negatives, via visual scenes that included pairs of matching and mismatching images for the direct object of the spoken sentence (smooth/wrinkled shirts). We believe that the factual/counterfactual contrast for the simple negatives would have been facilitated as well by the pronunciation of their stimuli; the simple sentences were uttered with prosodic focus on *has* or *hasn't*. Therefore, they established a binary situation similar to that of Kaup et al. (2006) and the binary contexts of Orenes et al. (2014). The results for the simple sentences indeed echoed those findings. For simple affirmatives, there was an increase in fixations to the image matching the mentioned state immediately following the verb. For simple negatives, fixations to the image matching the factual state increased more gradually.

For the cleft sentences, fixations to the target image rose during the critical time window at the same rate for affirmative and negative sentences, demonstrating that not all negative sentences automatically produce processing delays. These latter results are incompatible with the most straightforward predictions of two-stage models (e.g., Carpenter & Just, 1975; Kaup et al.,

2006, 2007). In addition, the affirmative clefts exhibited less rapid increases in fixations to the target than the affirmative simple sentences. Tian et al. argued that this was due to the greater pragmatic processing involved in the cleft sentences, such as the cleft potentially leading to a conversational implicature that would not be as readily generated with the simple affirmative sentences. This lends support to the prospect of our intonationally-driven implicature requiring relatively long processing times, at least when the Contrastive sentences are presented in a null discourse context.

In summary, prior studies of affirmative versus negative sentences have identified that these sentence types can produce distinct patterns of activation from each other for mentioned versus alternative states, especially when the predicate or discourse context facilitates a binary opposition. In such binary situations, the results suggest that comprehenders take longer to reach the final interpretation of negative sentences than affirmative ones. Explanations for this effect have included the need for negative sentences to involve additional perception-based simulations (e.g., Kaup et al., 2007), symbolic computations (e.g., Carpenter & Just, 1975), or pragmatic inferences (e.g., Tian et al., 2016) compared to affirmative sentences. However, pragmatic factors can modulate how negative sentences are processed, and so pragmatic processes must be at least part of the explanation.

Our target intonational contour can be used to evoke the negation of an alternative proposition, and like explicit negation involves a contrast in propositions. A negated sentence typically presupposes the question of whether the counterfactual state holds and asserts the factual one; our Contrastive sentences assert the literal state as holding in the past and implicate a change to an alternative state. Both sentence types have been argued to involve a temporal sequence of activation for the relevant alternatives. Compared to studies on the processing of

explicit negation, however, little online investigation had taken place on intonationally-driven implicature. We therefore drew on studies of explicit negation to look at how the activation of an intonationally-implicated state builds over time, and evaluate the explicit versus implicit negation of alternatives. To do so, we followed Kaup et al.'s (2006) design in using a picture-naming task with affirmative and negative sentences and multiple ISIs, but extended it by employing spoken sentences (in English), a larger set of ISIs, and the Contrastive Sentence Type. To our knowledge, no prior work has investigated intonationally-driven implicature with the particular configuration of Experiment 2. We review two studies that tested our target tune in contexts that highlighted alternatives, and then turn to the presentation of Experiment 2.

*Visual world studies of L+H* L-H% sentences*

Dennison (2010, Exp. 3) and Kurumada et al. (2014a) each used a visual-world experiment to examine the online processing of implicature for L+H* L-H% tunes, although with somewhat different research goals from each other and the present study. Dennison tested sentences such as *Lisa HAD the bell* pronounced with an L+H* L-H% contour or an L+H* L-L% contour (among other conditions, including affirmative and negative sentences with broad-focus intonation). A related study (Dennison, 2010, Exp. 1B) had shown that the L+H* L-L% contour allowed either an emphatic interpretation (that Lisa indeed had the bell) or (less strongly) the implicature that Lisa no longer has the bell. Dennison's design allowed an examination of whether the informational state at the point of the L+H* accent (i.e., prior to receiving the sentence-final rise or fall) would result in looks consistent with the implicature, and also allowed an investigation of the contribution of the final rise or fall to the pattern of looks.

As in the experiments presented here, Dennison was interested in how strongly the targeted implicature would emerge in a task that presented a range of felicitous[4] intonational contours. Unlike the present study, she established a situation in which the preceding context provided contrasting alternatives (objects were in one of two rooms, and the question under discussion was whether they had been moved or not). Critically, Dennison's situation supported either the emphatic interpretation or the implicature that the object had been moved. The results showed that listeners exhibited gaze patterns consistent with the implicated change of state only after the participants had processed the final rise or fall, and that such gaze patterns initiated earlier for sentences pronounced with the L+H* L-H% contour than with the L+H* L-L% one.

Kurumada et al. (2014a) also investigated whether listeners would generate an implicature upon hearing an accented verb, and also used an L+H* L-H% tune and a preceding context that provided alternatives (e.g., a visual scene including a zebra and an animal similar in appearance to a zebra, for *It LOOKS like a zebra*). However, in Kurumada et al.'s stimuli the L+H* accent always co-occurred with the phrase *It LOOKS like,* and was always followed by a final rise (among other differences from Dennison's study, such as the use of less complex visual scenes). None of the trials reinforced an emphatic interpretation of an L+H* accent. Notice as well that both the mentioned animal and its alternative were consistent with the literal assertion of the critical sentences. In this more supportive context, Kurumada et al. found that listeners directed fixations to the implicated alternative as soon as they had heard the accented verb. These results underline the importance of lexical, situational, or contextual support for pragmatic enrichment.

Taken together, the two visual world studies support the view that listeners often generate implicatures in response to an L+H* L-H% contour, and suggest that it is worthwhile to explore

how the processing of these sentences varies across different types of discourse situations. We now turn to our test of how quickly enriched meanings would be generated for sentences with the critical tune in the absence of a supportive context, and how the online processing of these sentences compares to ones with explicit negation.

*Participants*

Participants were recruited from the University of Hawai‘i at Mānoa community and randomly assigned to one ISI version of the experiment. Each participant received either course credit or five dollars as compensation and gave informed consent prior to participation. All participants were native speakers of English who had not learned any language other than English before age five. A total of 333 people participated: 54 participants per ISI for ISIs at 0, 500, 1000, 1500, and 2000 ms, 36 participants at an ISI of 2500 ms, plus an additional 27 participants across all ISIs who were subsequently eliminated because of poor performance according to the following criteria: voice key or naming errors resulting in less than three observations per test condition (10 participants), mean naming times exceeding 2.5 standard deviations from the grand mean across all participants (7 participants), or comprehension question accuracy lower than 80% (10 participants).

*Materials and design*

*Critical stimuli.* Experiment 2 tested thirty critical sentences from Experiment 1, using the same tokens of each of the three sentence types previously tested: Contrastive (L+H* L-H%), Affirmative, and Negative. The final rise of the Contrastive sentences began an average of 260 ms before the offset of the sentence, providing ample time for listeners to perceive the rise as the

predicate adjective unfolded. Sentence Type was crossed with two types of images. For critical stimuli, these always depicted the subject entity mentioned in the sentence. Table 4 presents the resulting 6 stimuli conditions. For Contrastive and Affirmative sentences, Image 1 represented the state mentioned in the sentences (e.g., a full mailbox for *The mailbox was full*). For the Negative sentences, it represented the state that was negated in the sentence, henceforth described as the *counterfactual-state* image. Image 2 depicted an opposite state (e.g., an empty mailbox); for the Negative sentences this was a *factual-state* image. These six conditions were counterbalanced within participants and items for each ISI in a Latin square design.

INSERT TABLE 4 ABOUT HERE

To ensure that there would be high name agreement for the critical pictures and similar naming times for the two image types, 64 critical images (32 pairs) were pretested. One image of each pair was assigned to one of two counterbalanced lists and presented to seventeen native English speakers randomly assigned to one of the two lists. Each image was presented in the center of a computer screen. The participant spoke the name of the depicted entity out loud as quickly as possible into a microphone, and the naming onset time was automatically recorded via a response box. Naming an image removed it from the screen and initiated the next trial. The 30 pairs subsequently selected for the main experiment were named with an average of 88% agreement to the preferred label, which consisted of one or two syllables for each item and always carried initial lexical stress. Naming times for these 30 selected pairs were converted to log RTs, and trimmed to remove any scores greater than three standard deviations from each participant's mean. A linear mixed-effects model testing the fixed effect of Image Type with

maximal random effects structure verified that the difference in naming times between the two image types (means of 860 and 863 ms) was not significant ($t = 0.26$).

*Filler stimuli.* Filler stimuli consisted of 70 trials in which the image to be named was not mentioned in the preceding sentence. The image was phonologically or semantically related to a mentioned word in 20 fillers, and unrelated to the preceding sentence in the remaining 50 fillers. Twenty-four fillers had the same syntactic form as the Negative critical sentences and the remainder used the syntactic form of the Affirmative and Contrastive critical sentences, although each type used a broader range of predicate adjectives (including many non-binary ones like *minty, red, dangling, spiky,* or *plastic*) to reduce the likelihood that participants would expect a predicate of a particular form. The fillers were recorded with a diverse set of intonational contours so that the contrastive contour in the critical sentences would not attract attention, and the L+H* accent would not predict a final rise. Thirty-five fillers received broad focus with a final fall, while the remaining 35 fillers received varied types of contours, including contrastive focus on the subject, verum focus with a final fall, question intonation (creating declarative questions), and L* L-H% tunes with unaccented *was*. Each sentence type of the critical trials was matched with a comparable sentence type in the fillers, so that, for example, the L+H* L-H% tune did not predict whether the upcoming image had been mentioned or not.

*Inter-stimulus intervals.* Each of the six stimulus conditions was tested at six ISIs, which ranged from 0 to 2500 ms in steps of 500 ms, chosen to fully capture the anticipated range of activation time for the set of interpretations. ISI was implemented as a between-participants factor, to avoid disruption to participants' behavior (Kaup et al., 2006).

*Procedure*

The experiment was conducted using E-Prime Version 1.2 (Psychology Software Tools) and a PST response box connected to a computer and a handheld microphone. Each participant was first seated in front of a computer to practice activation of the voice key. Once the voice key was properly adjusted in sensitivity, the participant received four randomly ordered practice trials, and then proceeded to the main experimental session.

On each trial, participants listened to a pre-recorded sentence while looking at a blank computer monitor. The sentence offset led to a predetermined ISI of 0 to 2500 ms, followed by the appearance of a picture at the center of the computer screen. Participants were told to verbally identify the depicted entity as quickly and accurately as possible. The speech onset was automatically recorded, triggering the immediate removal of the picture from the screen. After the naming component, comprehension questions appeared on the screen for 50% of the filler trials. Participants answered these aloud. Because the image for the critical trials always matched the subject of the sentence, 25 questions asked about the predicate (e.g., *How was the toothpaste?* after the sentence *The toothpaste was minty*), and the remaining 10 questions asked about the sentential subject (e.g., *What was yummy?* for *The doughnut was yummy*). All verbal responses to the pictures and questions were recorded into a digital voice recorder and were subsequently coded to check the accuracy of both picture naming and answers to the comprehension questions. Each experiment session lasted less than 25 minutes.

*Predictions*

Since we modeled our study on Kaup et al. (2006), we anticipated that a similar general pattern would hold for our results, namely that the relative activation for each Image Type would differ across Sentence Types and ISIs. It's useful to consider how the pragmatic situation of Experiment 2 differs from the other studies described above. In Experiment 2 the image always appeared after the spoken sentence, so there was no preceding visual context for any sentence. Because the content of the predicate was not predictable, listeners needed to recognize the predicate adjective to know what state to activate, negate, or use as the basis of an implicature. The average duration of the predicate adjectives varied by less than 20 ms across conditions (see Table 2 above), so the timing of critical information was well-controlled.

The critical stimuli had binary predicates, allowing easy activation of an opposite state. The binary-situation studies reviewed above suggest that Negative sentences should produce a match advantage for the factual image roughly 1000 to 1500 ms after the end of the sentence. However, all predictions for Experiment 2 should be treated with caution since Experiment 2 differed from the earlier work in a number of dimensions (e.g., spoken stimuli versus the written sentences of Kaup et al. (2006)). For the Affirmative sentences, the previous work suggests a match advantage at earlier ISIs than for the Negative sentences.

For the Contrastive sentences, we expected that the intonational information would introduce more complexity than the explicit negation of the Negative sentences. First, the listener must resolve which alternatives are being highlighted by the L+H* accent on *was*. A listener can take the alternatives to be a simple contrast between *was* and *was not*, and get an emphatic interpretation. Or, the listener can construct a set of alternatives and generate an implicature that the mentioned state no longer holds. Based on the results of Experiment 1, we predicted that listeners would prefer the implicature by a strong margin. We assume that this interpretation

requires the listener to update the mental model of the discourse to include both the asserted information and the implicated information (e.g., the mailbox was full at some point in the past, and also, it is no longer full). Because of the number of processing steps involved in the comprehension of the Contrastive sentences, we predicted that any effect of Image Type would emerge earlier for the Negative sentences than the Contrastive ones.

*Results*

Only trials in which the participant correctly named the picture and successfully triggered the voice key were included in the analysis; 6.3% of trials failed to reach this criterion. The reaction time distribution was positively skewed, as is common for reaction time scores, and so the data were converted to log naming times to more closely match a normal distribution. They were then treated for potential outliers by replacing all naming times greater than 2.5 standard deviations from the mean for each ISI with the cutoff value, affecting 1.8% of the trials. The resulting data were fit with a series of linear mixed-effects regression models using the *lme4* package (Bates, Maechler, Bolker, & Walker, 2015) in the statistical environment R (R Core Team, 2016). Preliminary analyses of the data early in data collection had indicated that all effects were diminished at the 2500 ms ISI, and so data collection was terminated for that ISI before a complete sample of participants had been tested. The analyses were therefore limited to ISIs of 0 to 2000 ms.

We conducted two sets of analyses. One set included all three factors (Sentence Type, Image Type, and ISI). The other set examined the relative naming time within each Sentence Type for each Image Type, across ISIs. That is, in these analyses each Sentence Type was modeled separately. In all analyses, ISI was treated as a continuous predictor; Sentence Type and

Image Type were categorical. Each predictor was scaled to a range of 1 and centered. We began the models with the maximal random effects structure for participants and items. Model comparison resulted in the removal of the Image Type slope from the random effects structure for items in each model; all other slopes remained. Significance levels were calculated using Satterthwaite approximations with the R package lmerTest (Kuznetsova, Brockhoff, & Christensen, 2016).

For ease of presentation, we provide brief discussion of the results of each model as we present the outcome of the model, and start with the models for each separate Sentence Type, which correspond to an analysis for each panel in Figure 2. Figure 2 displays the mean naming times in log ms for each Image Type and ISI, separated across panels by Sentence Type. The upper legend shows the color of each of the six Sentence Type by Image Type conditions. The lower legend indicates that all conditions that displayed Image 1 are graphed with solid lines, while those for Image 2 have dashed lines. Error bars indicate 95% confidence intervals for each mean, corrected for repeated measures.

INSERT FIGURE 2 ABOUT HERE

*Contrastive Sentence Type.* The top panel shows the results for the Contrastive Sentence Type, revealing that naming times remained relatively level across ISIs for the Contrastive Implicated condition (dashed line), while naming times for the Contrastive Literal condition tended to increase as the ISI increased (solid line). The statistical model for the Contrastive Sentence Type showed a marginal effect of ISI ($\beta = 0.046$, $t = 1.8$, $p < .10$). Image Type did not produce a significant main effect ($\beta = -0.010$, $t = -1.2$), but critically, there was a significant interaction of

Image Type and ISI (β = -0.069, $t$ = -2.8, $p$ < .01). This latter finding indicates a significant change between Contrastive Literal versus Contrastive Implicated conditions across ISIs.

*Affirmative Sentence Type.* The Affirmative Sentence Type produced a significant effect of ISI (β = 0.053, $t$ = 2.1, $p$ < .05), but not of Image Type (β = -0.009, $t$ = -1.1), nor their interaction (β = -0.029, $t$ = -1.2). Thus, with the Affirmative Sentence Type the two Image Types had similar naming times to each other at every ISI. Although this differed from Kaup et al. (2006), it is consistent with previous research suggesting rapid activation of both ends of a scale with binary predicates (Deese, 1964, 1965; Frazier et al., 2008; Gross et al., 1989; Kennedy & McNally, 2005), and also with Kaup et al.'s results for the late ISI.

*Negative Sentence Type.* The Negative Sentence Type displayed a third pattern: there was a significant main effect of Image Type (β = -0.031, $t$ = -3.7, $p$ < .01), but ISI was not significant (β = 0.042, $t$ = 1.6), nor was their interaction (β = 0.005, $t$ = 0.2). The Negative sentences therefore resulted in (numerically) shorter naming times at each ISI for the image associated with the asserted, factual meaning versus the image that matched the counterfactual meaning. This pattern was similar to what Kaup et al. found at their late ISI, and to the most similar conditions in Orenes et al. (2014) and Tian et al. (2016). The 95% confidence intervals shown in Figure 2 indicate that the difference between Image Types was most pronounced at the 1500 ms ISI, which is comparable to the time course found in previous studies.

*Affirmative-Contrastive model.* We evaluated effects of Sentence Type and its interactions with Image Type and ISI by creating mixed-effects models for each pair of Sentence Types. The Affirmative and Contrastive Sentence Types revealed quite similar naming times, differing most at the 2000-ms ISI. The Affirmative-Contrastive model produced a significant effect of ISI (β = 0.492, $t$ = 2.0, $p$ < .05), and a significant interaction of ISI and Image Type (β = -0.049, $t$ = -2.9,

$p < .01$), plus marginal effects of Sentence Type (β = -0.010, $t$ = -1.7, p < .10) and Image Type (β = -0.010, $t$ = -1.7, $p < .10$). No other predictor or interaction reached significance (all $|t|$'s < 1.3). The similarity of these Sentence Types was unexpected, especially considering that Experiment 1 produced very different continuation patterns for this pair of Sentence Types, using the same sentence tokens. We offer some potential explanations in the Discussion section below.

*Affirmative-Negative model.* The Negative Sentence Type had somewhat longer naming times in general compared to the Affirmative Sentence Type, with greater separation within ISIs between the naming times for Image 1 versus Image 2. This was reflected in a marginal interaction of Sentence Type and Image Type (β = -0.021, $t$ = -1.8, $p < .10$) in the Affirmative-Negative model. Image Type had a significant effect (β = -0.020, $t$ = -3.4, $p < .01$), driven by the strong effect of Image Type within the Negative Sentence Type. There was also a marginal effect of ISI (β = 0.048, $t$ = 1.9, $p < .10$). No other predictor or interaction reached significance (all $|t|$'s < 1).

*Negative-Contrastive model.* The Negative versus Contrastive Sentence Types revealed the most dramatic differences, with the Negative Counterfactual condition (black solid line) exhibiting relatively long naming times at each ISI, while its Contrastive counterpart (orange solid line) showed the greatest overall range and steepest slope across ISIs. Coupled with the relatively flat naming times across ISIs for the Contrastive Implicated condition (red dashed line) versus the gradually increasing times for the Negative Factual condition (gray dashed line), there was a significant three-way interaction among Sentence Type, ISI, and Image Type (β = -0.071, $t$ = -2.1, $p < .05$). The model also produced a significant main effect of Sentence Type (β = -0.015, $t$ = -2.5, $p < .05$) due to overall greater naming times for Negative than Contrastive sentences, and a significant main effect of Image Type (β = -0.021, $t$ = -3.3, $p < .01$). There were marginal effects of ISI (β = 0.045, $t$ = 1.8, $p < .10$), Sentence Type * Image Type (β = 0.019, $t$ = 1.7, $p <$

.10), and ISI * Image Type (β = -0.031, $t$ = -1.8, $p$ < .10). Sentence Type * ISI did not reach significance ($t$ < 0.4). These results are consistent with the prediction that Negative sentences would show an earlier effect of Image Type than Contrastive sentences.

*Summary*. Overall the results demonstrated effects of Image Type that differed by Sentence Type and ISI. Most notably, they showed a significant interaction of Image Type and ISI for the Contrastive sentences, a significant main effect of Image Type of the Negative sentences, and a significant three-way interaction of Image Type, ISI, and Sentence Type for Negative versus Contrastive sentences. The results therefore indicate a significant change in the meaning activation pattern for the Contrastive sentences. They further show that activation patterns for the Negative sentences differed from the pattern with Contrastive sentences, suggestive of listeners activating the factual state for Negative sentences more quickly than the implicated state for Contrastive sentences.

*Discussion*

The Contrastive, Affirmative, and Negative Sentence Types each produced somewhat different patterns of naming times, in which only the Contrastive Sentence Type resulted in a significant interaction of ISI and Image Type. Our discussion will first consider the most striking findings for each Sentence Type, and then turn to some broader implications of the results for processing mechanisms. For the Contrastive conditions, the early ISIs produced naming times very similar to those for the Affirmative conditions, but at late ISIs, the Contrastive conditions looked more like early-to-mid-ISI Negative conditions. This pattern of naming times supports the proposal that, in the null context tested here, listeners first activated the asserted state for the L+H* L-H% sentences, perhaps in conjunction with alternatives, and then settled on the

implicated one. The naming times suggest that the implicated state was highly activated at late ISIs, in conjunction with decay or suppression of the original, mentioned state. These results align with those of Chevalier et al. (2008), who found differences over time in the inferences generated for contrastively accented *or* in French. Our results differ from the visual-world findings of Kurumada et al. (2014a), who employed a discourse context that strongly supported the enriched interpretation of L+H* L-H% sentences and found gaze patterns consistent with a very rapid preference for that interpretation. They are more similar to the visual-world findings of Dennison (2010), whose contexts were relatively neutral between the supplied alternatives. Together the three studies indicate that responses to the L+H* L-H% tune vary with the particular configuration of pragmatic context, and that in less supportive discourse contexts resolution to an enriched interpretation emerges only at later time points. Yet critically, the L+H* L-H% tune can lead to a significant activation advantage for the implicated state over the basic state even when used in a null discourse context.

Given the results of Experiment 1, it is likely the case that the Contrastive sentences received a mix of interpretations, so that listeners did not always reach an enriched interpretation. This may explain why the Contrastive Sentence Type did not emerge as significantly different from the Affirmative one. Future work could further explore this question by providing a context that preferentially supports the enriched interpretation with the type of predicates we tested, such as by establishing a present-tense situation (e.g., *Any moment now, Chloe will arrive home. Her dog WAS groomed…*) and comparing Contrastive materials like the current ones to present-tense version of the Affirmative and Negative sentences. Future work on the L+H* L-H% tune could also benefit from testing other focus positions than the one tested here, along with other types of predicates. For example, a sentence such as *The food was fresh* with narrow focus on the

adjective and a final rise allows the implicature that the food was *merely* fresh, and perhaps not gourmet quality (Cummins & Rohde, 2015), but placing the contrastive accent on *was* implies that the food is *no longer* fresh. Predicates that support an individual-level interpretation (i.e., unchanging across time stages) draw out an emphatic/confirmatory reading; consider *Matt WAS born here… (and so he's eligible for the presidency).*

Turning to the Negative sentences, this Sentence Type produced a significant main effect of Image Type in the two-factor model, and a marginal interaction of Sentence Type and Image Type in the Affirmative-Negative model. These results are suggestive of a more general pattern of suppression for the counterfactual interpretation. The results echo previous research suggesting a gradual rise in relative activation for the factual state with binary-context negative sentences (e.g., Tian et al., 2016).

For the Affirmative sentences, the most noteworthy result was the absence of the strong match effect that has been seen in previous research. For example, Kaup et al. (2006) found shorter naming times for their equivalent of the Affirmative Mentioned condition compared to the Affirmative Opposite one at a 750-ms ISI. One reason for the lack of a match effect in our Experiment 2 could be the use of nuclear pitch accents on the binary predicate adjectives. The explicit accentuation of the adjectives in our study may have led to fast activation of associates of the predicate. It may also be the case that the two experiments set up different implicit contrasts across the stimuli. The present study used multiple intonational patterns in the fillers, providing multiple patterns of information structure across the experiment. Critical Affirmative sentences always received broad focus, with H* accents on the subject and predicate adjective but no accent on the verb. Kaup et al. (2006) and Anderson et al. (2010) presented written stimuli, and so readers were left to fill in their own, implicit prosody. Because all of their stimuli

(both critical and filler items) used either affirmative or negative versions of the same basic syntactic form, listeners may have tended to interpret the affirmative sentences as emphasizing the affirmative state, instead of having broad focus. This could have resulted in stronger activation for the mentioned (emphasized) state in their affirmative sentences versus ours. The simple affirmative sentences in Tian et al. (2016), which produced an early preference for looks to the matching image, were in fact pronounced with verum focus (e.g., *John HAS ironed his brother's shirt*). Future work that tests verum-focus affirmative sentences alongside broad-focus sentences and ones with narrow focus on the predicate could shed light on the processing that occurs with these different intonational forms.

*Processing mechanisms.* Because neither the Affirmative nor the Contrastive sentences resulted in significant differences between the two Image Types at early ISIs, it is difficult to draw strong conclusions about the exact processing steps involved in the Contrastive sentences. However, there were significant differences between the Contrastive sentences and the Negative sentences. These data cast doubt on simple, two-stage processing models in which Contrastive and Negative sentences each involve initial activation of the mentioned state of the predicate adjective and then shift at the same rate to activation of the opposite state (i.e., the implicated state for Contrastive sentences and the factual state for the Negative sentences). Minimally, different time courses are involved for the two types of sentences. Importantly, the differences in activation patterns between the two Sentence Types are not readily accounted for by a timing difference between when the negative operator appeared in the stimulus and when the critical intonational information appeared. The accented *was* of the Contrastive sentences and the *not* of the Negative sentences occurred at roughly the same point prior to sentence offset, so both sentences contained an expression associated with alternative propositions that was realized in the same

time frame. The final rise of the Contrastive sentences began less than 200 ms into the predicate adjective, as the segmental information necessary for lexical recognition was still unfolding. Moreover, there was no hint of an advantage for the Contrasted Implicated state over the Contrastive Literal state at the first three ISIs, whereas the effect of Image Type for the Negative sentences began to emerge at the earliest ISI, suggesting that any minor timing differences would be insufficient to account for the differences between the Negative and Contrastive sentences.

The results of Experiment 2 show a broad similarity to recent findings by Husband & Ferreira (2016), who explored patterns of activation across time in conjunction with L+H* accents. They found that these accents induced initial activation for contrastive and non-contrastive associates of the accented material, followed by a subsequent process of suppression for the less relevant material. Likewise, the results of Experiment 2 suggest rapid activation of associates of the predicative adjectives in the Affirmative sentences, and suppression in the Negative and Contrastive sentences of the ultimately less relevant state. If such a suppression process occurs, it could be associated with the selection of an event state that can be used as the basis (in normal discourse circumstances) of subsequent processing, such as anticipation of what will happen next. We speculate that the Negative and Contrastive sentences could differ in processing time courses in part because the counterfactual state of the Negative sentences would typically have little relevance to an updated event model or an anticipated discourse continuation, while in the Contrastive sentences the mentioned and the implicated states each have some bearing on the event model and on the upcoming discourse, although the implicated state should have the greatest relevance.

## Conclusion

We presented two experiments that tested pragmatic enrichment of sentences like *The mailbox WAS full* pronounced with an L+H* L-H% tune. Previous research on this topic has drawn heavily from metalinguistic judgments (researcher intuitions, rating tasks, and forced-choice tasks) and from tests in which the critical sentence was placed into a context that made the enriched interpretation salient. Such studies are beneficial in investigating whether comprehenders *can* achieve the target interpretation in a particular context, but they are less informative about how reliably and routinely listeners reach the interpretation when their attention is not drawn to it. By using a sentence continuation task in Experiment 1, we examined the proportion of interpretations with an implicated state contrast (was full; is now empty) for sentences with L+H* L-H% intonation compared to affirmative and negative counterparts with broad-focus intonation. The results demonstrated that the rate of generating the implicated contrast was significantly higher with the Contrastive Sentence Type. However, it was not the only interpretation evidenced by the continuations; in fact, only 63% of the continuations in the Contrastive condition indicated the target implicature. The results thus add to the view that intonation regularly occurs in a many-to-many relationship with meaning.

Experiment 2 tested the same Affirmative, Negative, and Contrastive sentence tokens in a picture-naming task to see how the activation for mentioned versus alternative interpretations rose or fell across time. To our knowledge, it was the first study to have looked at the online processing of intonationally implicated alternative states versus explicit negation. Importantly, the experiment tested stimuli presented without a preceding linguistic or visual context, and made use of a rich variety of intonational contours on the filler stimuli so that the critical intonational tune would not be highlighted. Such conditions allow a more conservative test of the contribution of intonation to the construction of meaning than experiments that use a context that

establishes salient contrast prior to the critical sentence, or experiments with fillers that reinforce a limited set of interpretations.

The results for the Negative sentences largely replicated previous findings for binary-context negated sentences while extending the findings to a different task configuration. The Affirmative sentences displayed somewhat unexpected results, suggesting a need to further explore how different patterns of prosodic focus affect the activation of presuppositions or alternatives and the online construction of meaning. Sentences with an L+H* L-H% contour showed similarity to broad-focus affirmative sentences early in processing, and to broad-focus negated sentences at later stages. Lexical negation produced an earlier contrast in activation levels between mentioned and alternative states than the L+H* L-H% contour, even though both involve a contrast between two states.

Overall the results support and extend previous findings exploring the processing of negation and intonation. Together with previous research they demonstrate that an enriched meaning is common, but not ubiquitous, for L+H* L-H% sentences; the contour is neither necessary nor sufficient for the generation of conversational implicatures. They supplement the picture of how sentence processing differs across discourse contexts by providing findings from null-context spoken sentences with well-controlled prosodic form, and set the stage for additional tests of how affirmative, negative, and final-rise sentences with various focus patterns and predicate types are processed. They also suggest the need for more research investigating how intonational preferences shift across different speech situations or categories of speakers, how multiple cues work together to establish the preferred interpretation, and the specific processing steps involved in comprehending intonationally-boosted implicature.

**Funding**

This work was supported by the National Science Foundation [grant number 0921696].

**Acknowledgements**

We thank Barbara Kaup for sharing the stimuli from Kaup et al. (2006), and Victoria Anderson for producing the spoken stimuli and providing many other valuable contributions to this work. Matthew Dennison assisted with coding the continuations in Experiment 1. We are grateful for helpful advice from the editors and anonymous reviewers of this paper, audience members at conference presentations of this work in 2008 and 2009, and the members of Dennison's dissertation committee.

**Notes**

1 A related study, designed and run subsequently to this one, also found high rates of state contrast interpretations for a different set of L+H* L-H% sentences, tested among affirmative sentences with other intonational contours (Dennison, 2010, Exp 1B).

2 Kaup et al. (2007) pointed out that their test sentences (e.g., *The eagle was not in the sky*) provided little information about the actual state of affairs. Tian et al. (2010)'s simple negated events (e.g., *Jane didn't cook the spaghetti*) allowed an inference about the state of the direct object, yet still left open many possibilities about what might have occurred, including inferences about other activities the agent might have participated in (see also discussion by Tian et al., (2016)).

3 They seem to assume that the simple affirmative could simply address the question of what happened. However, their stimuli were pronounced with emphasis on *has*, which to us suggests a less general QUD.

4 Dennison labeled a pair of her conditions as the Emphatic True versus Emphatic False conditions. Each employed L+H* L-L% prosody. In the Emphatic True condition, the visual scene matched the emphatic assertion but not the implicated change of state. In the Emphatic False condition, the visual scene matched the implicated change. As described in the main text, Dennison (2010)'s Experiment 1B found that naïve participants reached each of these interpretations with L+H* L-L% tunes. Thus, the labels True and False apply to the visual stimuli with respect to the emphatic interpretation, and do not indicate overall felicity of intonational forms. The conditions could also have been given labels such as "Rise-Fall Emphatic Match" and "Rise-Fall Implicated Match".

**References**


Anderson, S. E., Huette, S., Matlock, T., & Spivey, M. J. (2010). Comprehending negated sentences with binary states and locations. In *Proceedings of the 32nd Annual Conference of the Cognitive Science Society. Mahwah, NJ: Lawrence Erlbaum Associates, pp. 1192-1197*.

Arnold, J. E. (2001). The effects of thematic roles on pronoun use and frequency of reference. *Discourse Processes, 31*, 137-162.

Arvaniti, A., & Garding, G. (2007). Dialectal variation in the rising accents of American English. In J. Cole & J. H. Hualde (Eds.), *Laboratory Phonology 9* (pp. 547–576). New York: Mouton de Gruyter.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67,* 1-48. doi:10.18637/jss.v067.i01.

Beckman, M. E., & Ayers, G. (1997). *Guidelines for ToBI labelling* (Vol. 3.0). Ohio: Ohio State University Research Foundation.

Bing, J. (1979). Aspects of English intonation. *Distributed by Indiana University Linguistics Club*.

Bolinger, D. (1958). A theory of pitch accent in English. *Word, 14*, 109-149.

Bott, L., & Noveck, I. (2004). Some utterances are underinformative: The onset and time course of scalar inferences. *Journal of Memory and Language, 51*, 437–457.

Bott, L., Bailey, T. M., & Grodner, D. (2012). Distinguishing speed from accuracy in scalar implicatures. *Journal of Memory and Language, 66,* 123-142.

Büring, D. (2007). Semantics, intonation and information structure. In: G. Ramchand & C. Reiss (Eds.) *The Oxford Handbook of Linguistic Interfaces*, *445-474*.

Carpenter, P. A., & Just, M. A. (1975). Sentence comprehension: a psycholinguistic processing model of verification. *Psychological Review, 82*, 45–73.

Chevallier, C., Noveck, I. A., Nazir, T., Bott, L., Lanzetti, V., & Sperber, D. (2008). Making disjunctions exclusive. *The Quarterly Journal of Experimental Psychology*, *61*(11), 1741-1760.

Clark, H. H., & Chase, W. G. (1972). On the process of comparing sentences against pictures. *Cognitive Psychology, 3*, 472–517.

Constant, N. (2012). English rise-fall-rise: A study in the semantics and pragmatics of intonation. *Linguistics and Philosophy*, *35*(5), 407-442.

Cummins, C., & Rohde, H. (2015). Evoking context with contrastive stress: effects on pragmatic enrichment. *Frontiers in Psychology*, *6*.

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.

Deese, J. (1964). The associative structure of some English adjectives. *Journal of Verbal Learning and Verbal Behavior, 3*, 347-357.

Degen, J., & Tanenhaus, M. K. (2015). Availability of alternatives and the processing of scalar implicatures: A visual world eye-tracking study. *Cognitive Science*, 1–30.

Dennison, H. Y. (2010). *Processing Implied Meaning through Contrastive Prosody*. Doctoral dissertation. University of Hawai'i at Manoa.

Dennison, H. Y., & Schafer, A. J. (2010). Online construction of implicature through contrastive prosody. In *Proceedings of the Conference on Speech Prosody Conference*. Chicago, USA.

De Villiers, J. G., & Flusberg, H. B. T. (1975). Some facts one simply cannot deny. *Journal of Child Language, 2*, 279-286.

Frazier, L., Clifton, C. J., & Stolterfoht, B. (2008). Scale structure: Processing minimum standard and maximum standard scalar adjectives. *Cognition, 106*, 299-324.

Glenberg, A. M., Robertson, D. A., Jansen, J. L., & Johnson-Glenberg, M. C. (1999). Not propositions. *Journal of Cognitive Systems Research 1*, 19-33.

Grice, H. P. (1975). Logic and conversation. *Syntax and Semantics, 3,* 41–58.

Gross, D., Fischer, U., & Miller, G. A. (1989). The organization of adjectival meanings. *Journal of Memory and Language, 28*, 92-106.

Gussenhoven, C. (2004). *The phonology of tone and intonation*. Cambridge, UK: Cambridge University Press.

Heeren, W. F., Bibyk, S. A., Gunlogson, C., & Tanenhaus, M. K. (2015). Asking or Telling – Real-time Processing of Prosodically Distinguished Questions and Statements. *Language and Speech*, 1–28.

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

Huang, Y. T., & Snedeker, J. (2009). Online interpretation of scalar quantifiers: Insight into the semantics–pragmatics interface. *Cognitive Psychology*, *58*, 376-415.

Huang, Y. T., & Snedeker, J. (2011). Logic and conversation revisited: Evidence for a division between semantic and pragmatic content in real-time language comprehension. *Language and Cognitive Processes*, *26*, 1161-1172.

Husband, E. M., & Ferreira, F. (2016). The role of selection in the comprehension of focus alternatives. *Language, Cognition and Neuroscience, 31,* 217-235.

Israel, M. (2004). The pragmatics of polarity. In L. R. Horn & G. Ward (Eds.), *The handbook of pragmatics* (pp. 701-723). Malden, MA: Blackwell.

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

Jackendoff, R. (1972). *Semantic interpretation in generative grammar*. Cambridge, MA: MIT press.

Johnson-Laird, P. N. (2001). Mental models and deduction. *Trends in Cognitive Sciences, 5,* 434–442.

Kaup, B., Yaxley, R. H., Madden, C. J., Zwaan, R. A., & Lüdtke, J. (2007). Experiential simulations of negated text information. *The Quarterly Journal of Experimental Psychology, 60*, 976-990.

Kaup, B., Lüdtke, J., & Zwaan, R. A. (2006). Processing negated sentences with contradictory predicates: Is a door that is not open mentally closed? *Journal of Pragmatics, 38*, 1033-1050.

Kehler, A., Kertz, L., Rohde, H., & Elman, J. L. (2008). Coherence and coreference revisited. *Journal of Semantics*, *25*, 1-44.

Kennedy, C., & McNally, L. (2005). Scale structure, degree modification, and the semantics of gradable predicates. *Language, 81*, 345-381.

Kjelgaard, M. M., & Speer, S. R. (1999). Prosodic facilitation and interference in the resolution of temporary syntactic closure ambiguity. *Journal of Memory and Language*, *40*, 153-194.

Kurumada, C., Brown, M., Bibyk, S., Pontillo, D. F., & Tanenhaus, M. K. (2014a). Is it or isn't it: Listeners make rapid use of prosody to infer speaker meanings. *Cognition, 133*, 335–342.

Kurumada, C., Brown, M., Bibyk, S., Pontillo, D., & Tanenhaus, M. K. (2014b). Rapid adaptation in online pragmatic interpretation of contrastive prosody. *Proceedings of the 37th Annual Meeting of the Cognitive Science Society* (pp. 791-796). Austin, TX: Cognitive Science Society.

Kurumada, C., Brown, M., & Tanenhaus, M. K. (2012). Pragmatic interpretation of contrastive prosody; It *looks like* speech adaptation. In N. Miyake, D. Peebles, & R. P. Cooper (Eds.), *Proceedings of the 34th Annual Conference of the Cognitive Science Society* (pp. 647–652). Austin, TX: Cognitive Science Society.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2016). lmerTest: Tests in linear mixed effects models. R package version 2.0-32. http://CRAN.R-project.org/package=lmerTest

Ladd, D. R. (1983). Phonological features of intonational peaks. *Language,* 59, 721–759.

Ladd, D. R. (2008). *Intonational phonology*. Cambridge, UK: Cambridge University Press.

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

Lee, C. (2000). Contrastive predicates and conventional scales. In A. Okrent & J. Boyle (Eds.) *Proceedings from the Annual Meeting of the Chicago Linguistic Society, 36,* 243-257.

Levinson, S. C. (2000). Presumptive meanings: The theory of generalized conversational implicature. Cambridge: MIT Press.

Lüdtke, J., & Kaup, B. (2006). Context effects when reading negative and affirmative sentences. In *Proceedings of the 28th Annual Conference of the Cognitive Science Society* (pp. 1735-1740). Mahwah, NJ: Lawrence Erlbaum Associates.

Nieuwland, M. S., & Kuperberg, G. R. (2008). When the truth is not too hard to handle an event-related potential study on the pragmatics of negation. *Psychological Science*, *19*, 1213-1218.

Orenes, I., Beltrán, D., & Santamaría, C. (2014). How negation is understood: Evidence from the visual world paradigm. *Journal of Memory and Language*, *74*, 36-45.

Pierrehumbert, J., & Hirschberg, J. (1990). The meaning of intonational contours in interpretation of discourse. In P. Cohen, J. Morgan, & M. Pollack (Eds.), *Intentions in Communication*. Cambridge, USA: MIT Press.

Pierrehumbert, J. B., & Steele, S. A. (1989). Categories of tonal alignment in English. *Phonetica*, *46*, 181-196.

R Core Team (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Roberts, C. (1996/2012). Information structure in discourse: Towards an integrated formal theory of pragmatics. *Semantics & Pragmatics, 5,* 1-69.

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

Steedman, M. (2000). The syntactic process. Cambridge, USA: MIT Press.

Tian, Y. (2014). Negation processing: A dynamic pragmatic account. Doctoral dissertation, University College London.

Tian, Y., Breheny, R., & Ferguson, H. J. (2010). Why we simulate negated information: A dynamic pragmatic account. *The Quarterly Journal of Experimental Psychology*, *63*, 2305-2312.

Tian, Y., Ferguson, H., & Breheny, R. (2016). Processing negation without context–why and when we represent the positive argument. *Language, Cognition and Neuroscience*, *31*, 683-698.

Trabasso, T., Rollins, H., & Shaughnessy, E. (1971). Storage and verification stages in processing concepts. *Cognitive Psychology, 2*, 239–289.

Veilleux, N., Shattuck-Hufnagel, S., & Brugos, A. (2006). 6.911 Transcribing Prosodic Structure of Spoken Utterances with ToBI. from MIT Open Courseware: http://ocw.mit.edu/courses/electrical-engineering-and-computer-science/6-911-transcribing-prosodic-structure-of-spoken-utterances-with-tobi-january-iap-2006/index.htm

Wagner, M. (2012). Contrastive topics decomposed. *Semantics & Pragmatics, 5,* 1-54.

Ward, G., & Hirschberg, J. (1985). Implicating uncertainty: The pragmatics of fall-rise intonation. *Language*, 747-776.

Ward, G., & Hirschberg, J. (1988). Intonation and propositional attitude: The pragmatics of L*+ H LH%. In *Proceedings of the Fifth Eastern States Conference on Linguistics* (pp. 512-522). Ohio State University Press Columbus, OH.

Warren, P. (2016). *Uptalk: The phenomenon of rising intonation*. Cambridge University Press.

Zwaan, R. A., Stanfield, R. A., & Yaxley, R. H. (2002). Language comprehenders mentally represent the shapes of objects. *Psychological Science, 13*, 168-171.

## Appendix A: Critical stimuli for Experiment 1 and 2

*Critical items (starred items were not included in Experiment 2)*

1. The baby was dressed.
2. His beard was long.
3. The bed was made.
4. The belt was buckled.
5. The cage was locked.
6. The candle was lit.
7. The car was new.
8. The cat was fat.
9. The coat was buttoned.
10. The curtain was closed.
11. The dog was groomed.
12. The door was ajar.
13. The drawer was shut.
14. The dress was loose.*
15. The gift was wrapped.*
16. The jar was full.
17. The leg was hairy.
18. The mailbox was full.
19. The necklace was latched.
20. The pants were folded.
21. The pencil was sharp.
22. The plate was clean.
23. The road was curvy.
24. The rope was coiled.
25. The rug was rolled up.
26. The shirt was wrinkly.
27. The shoe was tied.
28. The stocking was hung.*
29. The suitcase was open.
30. The tire was flat.
31. The tree was leafy.
32. The window was broken.
33. The zipper was done.

## Appendix B: Acoustic analysis of critical stimuli

Mean duration and pitch excursions values for the critical stimuli appear in Tables 2 and 3, respectively of the main text. One-way ANOVA tests for each sentential region found that the mean duration differed significantly across the conditions as expected, since they differed in

accentuation: subject, $F(2, 64) = 10.39$, $p < .01$; auxiliary, $F(2, 64) = 348.59$, $p < .01$; and predicate, $F(2,64) = 3.47$, $p < .05$. The accented *WAS* in the Contrastive condition (C1) was longer than the unaccented auxiliary phrases of the other two conditions (C1−C2 = 145 ms, $SE = .007$, $p < .01$; C1−C3 = 156 ms, $SE = .008$, $p < .01$). The accented predicate adjectives in the Affirmative (C2) and Negative (C3) conditions were likewise longer than the non-accented predicate adjective in the Contrastive condition (C2−C1 = 19 ms, $SE = .008$, $p < .05$; C3−C1 = 17 ms, $SE = .008$, $p < .05$). However, the pre-nuclear accented subject in C3 was significantly shorter than both the unaccented subject in C1 and the pre-nuclear accented subject in C2 (C3−C1 = −26 ms, $SE = .009$; $p < .01$ C2−C1 = −38 ms, $p < .01$, $SE = .007$, $p < .01$), presumably due to the presence of an additional word (*not*) in this condition and the speaker's attempt to maintain the timing and rhythm of her utterances to keep them as similar as possible across the conditions.

F0 values also showed differences across conditions for each sentential region. Table 3 of the main text presents the mean F0 values at six different points in the sound files. One-way repeated-measures ANOVA tests found significant differences across conditions for: F0 maximum values in the subject, $F(2, 64) = 94.02$, $p < .001$; F0 minimum values in the subject region following that maximum, corresponding to the fall for the L target in C1's L+H* on *WAS* versus the interpolation between two H* accents for C2 and C3, $F(2, 64) = 381.29$, $p < .001$; F0 maxima in the auxiliary region, $F(2, 64) = 157.5$, $p < .001$; and the final F0 at the end of the sentence, $F(2, 64) = 284.9$, $p < .001$. Table 3 additionally shows the average F0 minimum for the predicate in the Contrastive condition, and the average F0 maximum for the accented predicate adjective in the other two conditions.

Pairwise comparisons revealed additional effects. In the subject region, the F0 maximum was significantly lower for the non-accented subject in the Contrastive than the H*-accented subjects in the Affirmative ($C1-C2$ = −28.85 Hz, $SE$ = 2.3, $p < .001$) and Negative conditions ($C1-C3$ = −25.92 Hz, $SE$ = 2.8, $p < .001$). The after-peak F0 minima in the subject region was significantly lower for the Contrastive condition than the other two ($C1 – C2$ = –58.56 Hz, $SE$ = 2.55, $p < .001$; $C1 – C3$ = – 52.16 Hz, $SE$ = 2.54, $p < .001$), as well as lower in the Negative than the Affirmative condition ($C2 – C3$ = 6.4 Hz, $SE$ = 1.8, $p < .01$). F0 maxima in the auxiliary were greater with the L+H*-accented auxiliary in the Contrastive condition than the non-accented auxiliary in the other two conditions ($C1-C2$ = 40.4 Hz, $SE$ = 2.5, $p < .001$; $C1-C3$ = 52.9 Hz, $SE$ = 3.6, $p < .001$). The F0 maximum on the auxiliary was also greater in the Affirmative than in the Negative condition ($C2-C3$ = 12.5 Hz, $SE$ = 3.2, $p < .001$). In both of these conditions, this value represents the F0 that occurs between two H* pitch accents. Finally, F0 values at the end of the sentence confirmed that the Contrastive maximum was significantly higher than the maximum for each of the other two conditions and that there was no significant difference in the F0 maximum values between the two neutral conditions; only the Contrastive condition ended with a high tone ($C1-C2$ = 86.2 Hz, $SE$ = 4.9, $p < .001$; $C1-C3$ = 92.1 Hz, $SE$ = 4.2, $p < .001$). Thus, the acoustic measurements are consistent with the phonological descriptions of the tunes from the ToBI analysis.

**Tables**

Table 1. Three sentence conditions in Experiment 1. (Regular capital letters mark the L+H* accent whereas small capital letters indicate the H* accent.)

| | Condition | Example |
|---|---|---|
| C1 | Contrastive | L+H* L-H%<br>The pencil WAS sharp... |
| C2 | Affirmative | H* H* L-L%<br>The PENCIL was SHARP. |
| C3 | Negative | H* H* H* L-L%<br>The PENCIL was NOT SHARP. |

Table 2. Mean duration (in milliseconds) of each sentential region in each test condition in Experiment 1; the underlined numbers represent values from the accented items.

| Condition | Subject | Auxiliary | (*not*) | Predicate | Total |
|---|---|---|---|---|---|
| C1 (Contrastive) | 421 | <u>304</u> | - | 443 | 1168 |
| C2 (Affirmative) | <u>433</u> | 159 | - | <u>462</u> | 1054 |
| C3 (Negative) | <u>396</u> | 148 | <u>273</u> | <u>461</u> | 1278 |

Table 3. Mean F0 values (in Hz) at six key points in each test condition in Experiment 1.

| | Subject | | | Auxiliary | Predicate | |
|---|---|---|---|---|---|---|
| | Initial F0 | Maximum | Following min | Maximum | Min (C1) / Max (C2 & C3) | Final F0 |
| C1 | 188 | 196 | 139 | 251 | 139 (min) | 228 |
| C2 | 187 | 224 | 198 | 211 | 236 (max) | 141 |
| C3 | 185 | 222 | 191 | 198 | 232 (max) | 135 |

Table 4. Six experimental conditions crossing 3 Sentence types with 2 Image types

| Cond. | Sentence | | Image | | State of affairs |
|---|---|---|---|---|---|
| 1 | 1<br>Contrastive | L+H* L-H%<br>The mailbox WAS full. | 1 | Full mailbox | Mentioned state (Literal interpretation) |
| 2 | | | 2 | Empty mailbox | Opposite state (Implicated interpretation) |
| 3 | 2<br>Affirmative | H* H*L-L%<br>The MAILBOX was FULL. | 1 | Full mailbox | Mentioned state |
| 4 | | | 2 | Empty mailbox | Opposite state |
| 5 | 3<br>Negative | H* H* H*L-L%<br>The MAILBOX was NOT FULL. | 1 | Full mailbox | Counterfactual state |
| 6 | | | 2 | Empty mailbox | Factual state |

**Figure captions**

Figure 1. Sample F0 tracks for the test sentences in (a) Contrastive (b) Affirmative, and (c) Negative conditions, from top to bottom.

Figure 2. Picture naming times in log ms for (a) Contrastive, (b) Affirmative, and (c) Negative Sentence Types, from top to bottom, by Image Type and interstimulus interval (ISI) in ms. Error bars represent 95% confidence intervals of the mean.

Figure 1

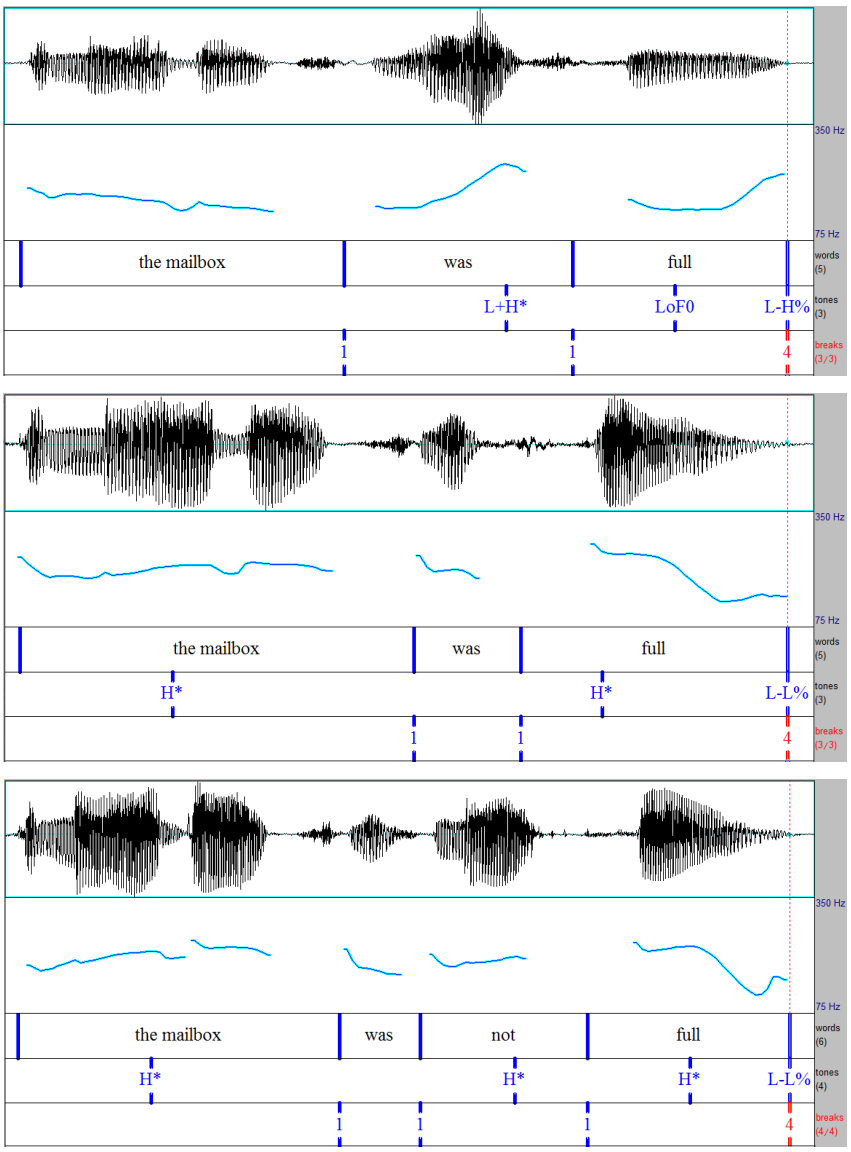
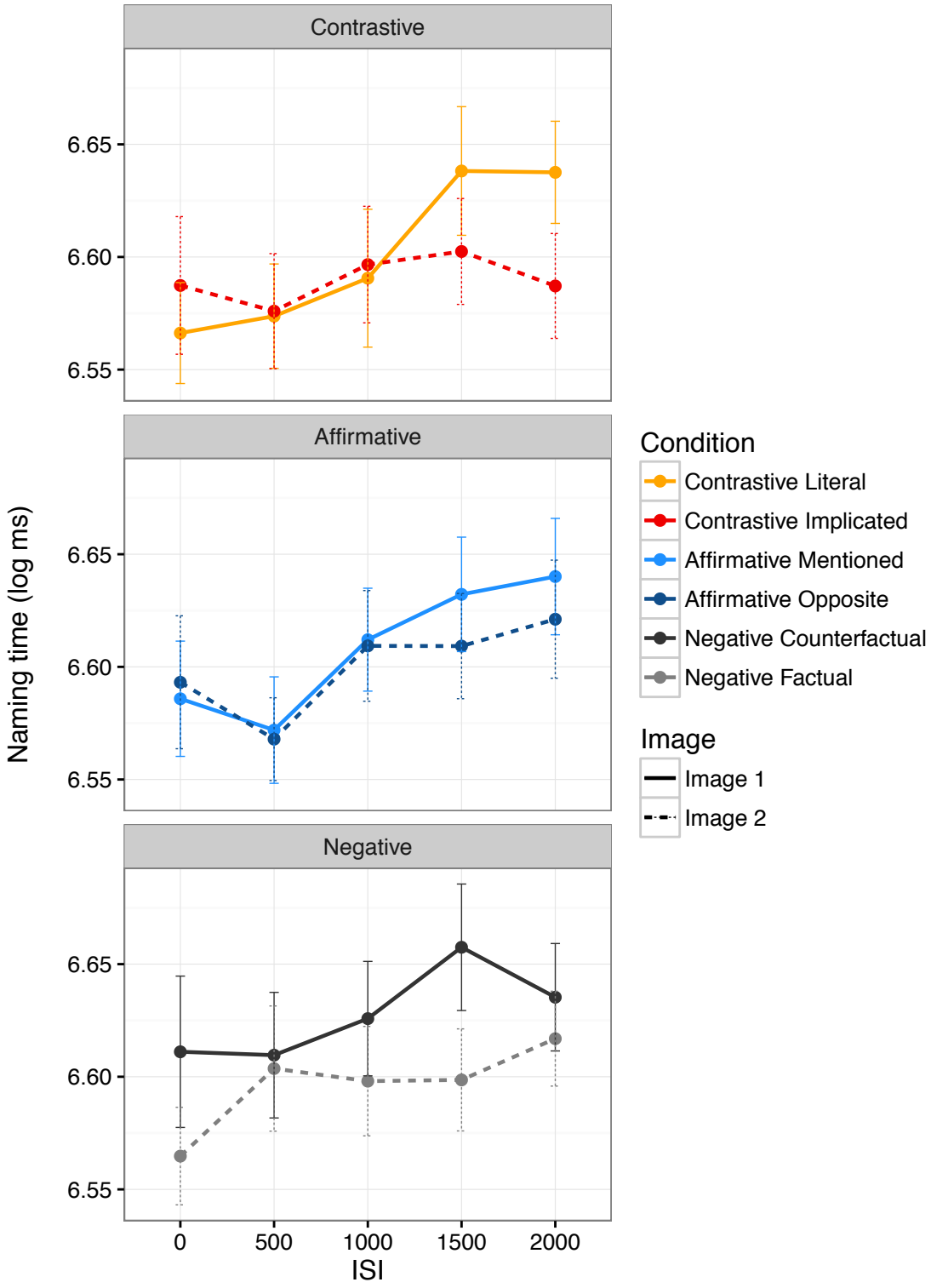

350 Hz
75 Hz
the mailbox
was
full
words (5)
L+H*
LoF0
L-H%
tones (3)
1
1
4
breaks (3/3)
350 Hz
75 Hz
the mailbox
was
full
words (5)
H*
H*
L-L%
tones (3)
1
1
4
breaks (3/3)
350 Hz
75 Hz
the mailbox
was
not
full
words (6)
H*
H*
H*
L-L%
tones (4)
1
1
1
4
breaks (4/4)

Figure 2


Contrastive
Affirmative
Negative
Naming time (log ms)
6.65
6.60
6.55
0
500
1000
1500
2000
ISI
Condition
Contrastive Literal
Contrastive Implicated
Affirmative Mentioned
Affirmative Opposite
Negative Counterfactual
Negative Factual
Image
Image 1
Image 2