Individual Differences in the Production and Processing of Focus Intonation

A thesis submitted in partial satisfaction of the requirements for the degree Master of Arts in Linguistics

by

Adam J. Royer

2017
ABSTRACT OF THE THESIS

Individual Differences in the Production and Processing of Focus Intonation

by

Adam J. Royer

Master of Arts in Linguistics
University of California, Los Angeles, 2017
Professor Sun-Ah Jun, Chair

The relationship between the perception and production of prosody is less well studied than the perception and production of segments. This study consists of two experiments to examine the relationship between production of focus intonation in English and perception of focus intonation in a visual world paradigm eye tracking task. The results suggest that there is no influence of perception on the production of focus intonation and no connection between Autism Quotient and the perception of focus intonation. The fact that very few subjects were helped about felicitous focus intonation suggests that the null results are the result of a confound.
The thesis of Adam J. Royer is approved.

Jesse A. Harris
Megha Sundara

Sun-Ah Jun, Committee Chair
Contents

1 Introduction 1

2 Background 3
   2.1 Intonational Phonology of English 3
   2.2 Meaning of H* versus L+H* 4
   2.3 Production and Perception Studies: Differences between H* and L+H* 5
   2.4 Individual Differences in Prosody 10
   2.5 Hypotheses and Predictions 13

3 Experiment 1: Production 15
   3.1 Participant information 15
   3.2 Materials 15
   3.3 Procedure 16
   3.4 Results 18

4 Experiment 2: Perception 24
   4.1 Visual and Auditory stimuli 24
   4.2 Eye tracking task 27
   4.3 Results 30
   4.4 Garden Pathing effect 37
   4.5 Post-Hoc Experiments 39
      4.5.1 Methods and Procedure 39
      4.5.2 Results and Discussion 40
5 Discussion and Conclusion

5.1 Individual Differences .................................................. 43
List of Figures

1. An example image of the ornament array from Ito and Speer (2008) .................. 8

2. The distribution of Autism Quotient scores of participants ($\bar{x} = 19, \sigma = 4.26$). AQ scores greater than 22 (approx $>+1$ SD above the mean) were considered high AQ, and scores lower than 16 (approx $<-1$ SD below the mean) were considered low AQ. 16

3. Block 1 non-contrastive trial. What was seen in the first half of the trial is on the left, and the second half of the trial appears on the right. ......................... 17

4. Block 1 contrastive trial. What was seen in the first half of the trial is on the left, and the second half of the trial appears on the right. ......................... 18

5. The distribution of pitch accents used by participants on the adjective (left) and on the noun (right) respectively in order to signal contrastive focus on the adjective. The ‘*?’ label is meant to indicate uncertainty about whether a word is accented or not. ................................................................. 19

6. The various tunes employed over the adjective-noun pair in order to mark focus on the adjective. ................................................................. 20

7. The various tunes employed over the adjective-noun pair, for various AQ scores, in order to mark focus on the adjective. ............................................. 22

8. Pitch contour and spectrogram of the instructions for the L+H* condition. .......... 25

9. Pitch contour and spectrogram of the instructions for the H* condition. .......... 26

10. Pitch contour and spectrogram of the first instruction set that created the discourse context. ................................................................. 26

11. An example of an array of objects used in the eye tracking task. ..................... 29

12. Log odds of looking to the target object cell for each condition. ...................... 33
13 Log odds of looking to the target object cell for each condition for high, mid, and low AQ groups. .................................................. 35
14 Log odds of looking to the target object cell for each condition for L+H* users and Non-L+H* users. .................................................. 36
15 Log odds of looking to the Target and Competitor object cells for Felicitous H* and Infelicitous L+H* conditions. .................................. 38
16 Log odds of looking to the Target and Competitor object cells for Felicitous L+H* and Infelicitous H* conditions. .................................. 38
17 Log odds of looking to the target object cell for each condition. .................. 40
18 Log odds of looking to the target object cell for each condition. .................. 41
List of Tables

1  The distribution of the proportion of pitch accents used, for various AQ scores, on
   the adjective in order to signal contrastive focus on the adjective.  . . . . . . . . . .  21

2  Comparisons of model fits for Autism Spectrum Quotient score as a predictor. . .  32

3  Estimates, standard errors, z values, and p values of best fitting model. . . . . . .  32
ACKNOWLEDGMENTS

I would like to thank the members of my committee, Sun-Ah Jun, Megha Sundara and Jesse Harris, for all of their support while working on this project. I would also like to thank the UCLA Phonetics Lab and Shari Speer for their valuable input. This project would not be possible if not for the hard work of my undergraduate research assistants, Michel Wong, Victoria “Vita” Khachaturyan, Neda Vesselinova, and Nigel Soriano. I am grateful to the UCLA Graduate Division for their funding of this project through a Graduate Summer Research Mentorship. All mistakes are my own.
1. Introduction

Prosody plays an important role in a speaker’s encoding, and a listener’s decoding, of intention and discourse meaning (Couper-Kuhle, 2001; Pierrehumbert and Hirschberg, 1990). One of the ways that these pragmatic meanings relate to prosody is through a speaker’s use of intonation. Intonation is the use of prosodic cues like f0, duration, and intensity, in order to convey post-lexical, sentence level meaning (Ladd, 2008). For example, a speaker’s use of a certain f0 excursion on the stressed syllable of a word can signal the information status of that word. In English for example, the H* pitch accent is proposed to be preferentially associated with new information items and words accented with a L+H* are associated with narrow focus. Through attending to the prosodic cues, a listener can decode discourse information or the intentions of the speaker. The pragmatic use of these prosodic cues, however, is not without some inter-speaker and intra-speaker variation. For example, Speer et al. (2011) found a high degree of variability in the way speakers used prosodic breaks to disambiguate ambiguous prepositional phrase attachment sentences. Despite this kind of variation, listeners are able to understand what a speaker is intending. Even between listeners, however, there is slight variation in the way they decode these meanings. The invariant prosodic cues related to a particular pragmatic meaning is what much of the work on production and perception of intonation has focused on. This type of question treats speaker variation as noise and does not address what systemic link is there between the variation we see in prosodic cues and individual variation in production and perception.

The purpose of this paper is to investigate the role of individual differences in the production and perception of contrastive focus intonation in English. This will be done by first laying out the background literature on English intonational phonology in the Autosegmental Metrical framework, as well as what work on individual differences in production and perception of prosody has already found. Next, in Sections 3 and 4, two experiments will be presented that examine the strategies
of individuals in producing and perceiving contrastive focus. The first experiment is a production study in which participants produced sentences which varied in the information status of target words (i.e., focused words and new information words). This allowed for phonological analysis of the intonation contour the individual used to mark the information status of the word. The second experiment was a Visual World Paradigm (VWP) eye tracking study using a visual search task. This experiment tested the participants’ sensitivity to the difference between two different pitch accent types (i.e., L+H* vs H*) in felicitous contrastive focus contexts and infelicitous contrastive focus contexts. The use of eye tracking allows for detailed time course information about a participant’s processing of various cues as segmental and suprasegmental information becomes available in the speech signal. Finally, the results will be discussed, in Section 5, in context of the individual differences of participants as well as future research directions for this study.
2. Background

2.1 Intonational Phonology of English

The framework of intonation adopted in this paper is the intonational phonology of English intonation in the autosegmental metrical framework. In the autosegmental metrical (AM) framework, an intonation contour is phonologically analyzed as a sequence of high (H) or low (L) tones and their combinations (Pierrehumbert, 1980; Beckman and Pierrehumbert, 1986; Ladd, 2008). The Mainstream American English Tones and Break Indices (MAE_ToBI) conventions for annotation of English intonation were used to analyze the production data in this study (Beckman and Hirschberg, 1994; Beckman and Ayers, 1997).

In the AM framework, tones mark either phrasing or prominence in an utterance. Phrasing is marked by f0 changes at a prosodic break as well as lengthening of the syllable adjacent to an initial or final boundary. In English, there are two levels of prosodic grouping; the intermediate phrase (ip) and the intonational phrase (IP). The intermediate phrase is marked by either a L- and H- phrase accent and intonational phrase breaks are marked by either a L% or H% boundary tone. Phrasing in English is closely associated with the syntactic structure of an utterance, however phonological constraints also play a role in phrasing.

Prominence is marked by ‘pitch accents’, which are associated with the stressed syllable of the accented word. In American English, there are five pitch accents; H*, L*, L+H*, L*+H, and H+!H*. The tone that is realized on the stressed syllable gets a star (e.g., T*) and the tone that precedes (i.e., leading tone) or follows (i.e., trailing tone) it in the immediately adjacent syllable is connected with a plus sign (e.g., T+ or +T). The use of the exclamation point is to indicate the lowering (i.e., downstepping) of a H tone relative to a previous H (e.g., !H*, L*+!H, H+!H*, etc.).

The pitch accents of focus in this paper are the H* and the L+H*. The H* accent is realized
as a gradual rise in f0 to a high target on the stressed syllable of the accented word. The L+H* accent is characterized by a steep rise in f0 from a low f0 target before the stressed syllable of the accented word, or at the onset of the stressed syllable, to a high f0 target toward the end of the stressed syllable. The high target is typically realized higher than the high target of a H* (Breen et al., 2010). This makes the L+H* perceptually very salient because of the steeper slope and higher f0 target relative to those of the H* (Ayers, 1996).

2.2 Meaning of H* versus L+H*

Pitch accent choice is one way in which speakers indicate the information status of an accented item. According to Pierrehumbert and Hirschberg (1990), the use of the H* pitch accent is to instantiate the accented item in the listener’s discourse model. In other words, the speaker is marking the accented word as “new” information which is to be added to the speaker and listener’s mutually shared beliefs. The example given, (1), is meant to indicate that with each H*, the speaker is introducing “files”, “deleted”, “deliberately” as information being added to the common ground of the speaker and listener.

In contrast, the use of a L+H* is to instantiate the accent word in the listener’s discourse model, to the exclusion of other salient alternatives. This accent serves to highlight the accented item as the focus of attention, to the exclusion of other items in a set of alternatives. The difference in meaning between the H* and the L+H* can be seen in (2), which is similar to (1) with only minor changes. The sentence in (2) could be preceded by the speaker saying, “You didn’t save my files...”, thereby introducing an alternative to “delete.” By using the L+H* pitch accent on “deleted”, the speaker in (2) intends to instantiate “deleted” in the common ground among plausible alternatives, such as “saved” or “moved”, which are not to be instantiated.
2.3 Production and Perception Studies: Differences between H* and L+H*

While H* and L+H* are proposed as separate pitch accent types by Beckman and Pierre-humbert (1986) and they have been proposed to have separate meanings by Pierrehumbert and Hirschberg (1990), it is still a matter of debate whether there really is a distinction between H* and L+H*. There has been experimental studies that have found that the proposed distinctions between H* and L+H* are only weakly supported by behavioral evidence. If these pitch accent types are not distinct, but rather allotones of the same pitch accent type, then the current model of the intonational phonology of English must be revised.

Using an identification task, Ladd and Morton (1997) found that while participants were successful in categorizing the two pitch accents into two distinct groups, there was not a clear perceptual boundary between “normal” (corresponding to H*) and “emphasized” (corresponding to L+H*) accents, but rather a large overlap in their categories. This suggests that the two accents are perceived in a gradient way. Also testing for phonetic gradience, Dilley (2010) found a similar result to Ladd and Morton. In Dilley’s study, participants were played stimuli words that were accented, where the peak f0 had been resynthesized in a stepwise manner. The pitch ranged from the f0 height of a typical H* to a higher f0 target that was typical of a L+H*. The height of the f0 on the syllable immediately preceding the high tone target was raised in 12 steps\(^1\). Dilley had participants imitate the various steps of the stimuli in order to see if participants produced categorical

\(^1\)Step 12 was a H* with a slow rise to the H target and Step 1 was a L+H* with a L target at the bottom of the speaker’s pitch range followed by a steep rise to the H target
or continuous differences in pitch height. If there is a categorical difference, the result would look like what Flege and Eefting (1988) found with VOT; speakers have a shift in production of VOT at a certain stimulus step which corresponds to a perceptual boundary. The results of Dilley’s study showed that participants do imitate the stimuli in a continuous way and do not exhibit this shift in production between a H* and a L+H*.

While previous studies have used identification and imitation tasks to test for the discreteness of H* and L+H*, recent research has taken advantage of on-line methods, like eye tracking, to investigate the interpretation of pitch accents accents. By continually sampling the sequences of fixations and saccades of the eye, researchers are provided with time-locked information about a participant’s processing of an utterance (Huettig et al., 2011). This time course information reveals the extent to which the processor uses certain cues and segmental/suprasegmental information for interpretation or prediction. Intonation is one cue that allows the processor to make predictions about upcoming material. It has been found in previous work that the processor is able to make immediate use of prosodic cues, as soon as they are available in the speech stream (Beach, 1991; Snedeker and Trueswell, 2003).

Watson et al. (2010) used eye tracking to test the interpretive domains of H* and L+H* and find differences in interpretation of the pitch accents. Using the VWP, they tested whether H* and L+H* are preferentially interpreted based on discourse status of the accented word (i.e., H* for words that are being provided as new information, L+H* for words that are in contrastive focus). The task involved participants moving objects on a monitor in relation to shapes that appeared in the corners of the screen. The array consisted of 4 objects and 4 shapes. Two of these objects were phonetic cohorts (e.g., “camel” and “candle”) and the other two were unrelated images (e.g., “dog” and “flower”).

In a trial, participants were first told to click on one of the cohorts and one of the unrelated images to set up a potential contrast (i.e., (3)). The cohort in the conjunct was then considered the “focus” cohort (e.g., camel), and the unmentioned cohort was considered the “new information” cohort (e.g., candle). The next set of instructions was to move the previously named unrelated
object (i.e., (4)) in order to set up the possibility of a contrast in the final set of instructions. The manipulations were made in the last set of instructions. First, either the discourse new or the focused object was named. The second manipulation was the type of pitch accent that was used on the cohort; either a H* (i.e., (5)) or a L+H* (i.e., (6)).

(3) **Instruction 1:** “Click on the camel and the dog.”

(4) **Instruction 2:** “Move the dog to the right of the square”

(5) **Instruction 3a, H*: “Now, move the camel\textsubscript{H}/candle\textsubscript{H} below the triangle”

(6) **Instruction 3b, L+H*: “Now, move the camel\textsubscript{L+H}/candle\textsubscript{L+H} below the triangle”

Watson et al. found that 200ms into the ambiguous onset and vowel of the nouns, participants looked towards both focus noun as well as the new information noun when they heard a H* on the noun. It takes about 200ms to plan and execute a saccade, so Watson et al. was investigating the looking patterns that happens between 200ms and 400ms after the onset of the noun. What’s critically important about this time window is that the segmental ambiguity persisted through the onset and vowel of the first syllable, so any eye movements made in this window would have to be made due to the intonation, not the segments. In this same time window, only looks to the focus noun increased when hearing a L+H* on the noun, while looks to the new information noun decreased. The H* pitch accent was interpreted felicitously in both discourse new and focus conditions, while the L+H* pitch accent was only interpreted felicitously in the focus condition. This finding blurs the distinction between H* and L+H* because of the overlap in interpretative domains. In other words, these two pitch accents are not in complementary distribution with regard to the discourse contexts in which they can appear felicitously.

One study that does support a distinction between H* and L+H*, however, was done by Ito and Speer (2008). Ito and Speer used eye tracking to examine the effect of pitch accent on predictions of upcoming items. In their study, participants were seated in front of an array of different colored tree ornaments (see Figure 1) while wearing a head mounted eye tracker. The array consisted
of 11 cells, where ornaments of the same type (e.g., bell, drum, etc.), but different color (e.g.,
orange, blue, etc.), were grouped together within a cell. Participants then heard sequences of two
instructions to find one of the ornaments on the board and place it on a miniature holiday tree and
then find a second ornament and hang it on the tree. The critical manipulation was in the intonation
of the second set of instructions. In all their experiments, participants heard adjective-noun pairs
(e.g., blue ball, brown drum, etc.) to uniquely pick out an object from the array in front of them.
In critical trials, the target word (sometimes the adjective, some times the noun depending on the
experiment) carried a L+H* pitch accent.

The use of the L+H* was meant to convey a focus interpretation on the adjective word. In a
sequence of instructions to hand various ornaments.

![Figure 1: An example image of the ornament array from Ito and Speer (2008)](image)

In the first experiment, the location of the L+H* accent was manipulated so that it was either
on the adjective or the noun, while keeping the segmental information the same. The noun in
the second set of instructions was always the same as the noun in the first set of instructions,
which created a context for contrasting the colors of the two ornaments, not contrasting the type of
ornament since they were repeated from the first set of instructions. This created a situation where
the L+H* was either used felicitously such as in (7) or infelicitously such as in (8) (the item with L+H* is capitalized).

(7) Hang the blue drum. Now, hang the \textsc{green}_{L+H^*} drum.

(8) Hang the blue drum. Now, hang the \textsc{green}_{L+H^*} drum.

What they found was that, in the adjective focused condition (i.e., (7)), there were anticipatory looks to the target ornament cell before the onset of the noun. That is to say that participants were looking at the target cell before they had any segmental information telling them what the target object was. Anticipatory looks were not found in the noun focused condition (i.e., (8)). The explanation for the anticipatory looks is that accenting the adjective with the L+H* caused the listener to consider the set of color alternatives of that same type of ornament, thereby constraining possible referents. In (8), just using a L+H* on the noun did not help participants to predict ornament because the noun was the last item in the phrase and evoking alternatives of the noun did not narrow down possible target cells.

In experiment 2, Ito and Speer tested whether infelicitous use of the L+H* in a non-contrastive context, see (9), would lead participants to incorrectly predict that the next type of ornament would be a repetition of the previous type of ornament mentioned in the first set of instructions. They also included a condition in which a neutral H* pitch accent was used on the adjective, as seen in (10). The addition of this neutral pitch accent condition was to examine the looking pattern when the intonation does not generate a set of alternatives.

(9) Hang the red angel. Now, hang the \textsc{green}_{L+H^*} drum.

(10) Hand the red angel. Now, hang the \textsc{green}_{H^*} drum.

The results of experiment 2 indicated that when participants heard the L+H* on the adjective, they mistakenly looked to the previously mentioned noun cell (i.e., they looked to the ‘angel’ cell instead of the target ‘drum’ cell in the case of (9)) before the onset of the noun, and later had to change their gaze to fixate in the target cell once they had heard and processed the noun. What this
shows is that the $L+H^*$ on the adjective led participants down a garden path where they predicted the upcoming noun to be the same as before.

There were no looks to the target object cell in the neutral intonation condition (i.e., (10)) until the segmental information of the noun was available. Participants did not use the $H^*$ on the adjective to predict the identity of the upcoming noun. This means that the $H^*$ did not evoke a set of alternatives that could be used to narrow down the possible ornament to be named. This is strong evidence that these two pitch accents have differing pragmatic uses.

2.4 Individual Differences in Prosody

Individual differences in production and perception of prosody have accounted for the differences in prosodic encoding and decoding. Focusing on production differences, Speer and Foltz (2015) conducted a cross modal priming study to investigate the role of implicit prosody on the priming of auditory stimuli. Participants were presented with visual primes (i.e., written text) that created a context that put the target word in focus. They then heard an auditory stimulus word and had to say whether the auditory stimulus matched the written target word.

The presence or absence of a pitch accent on the auditory stimulus word (i.e., $L+H^*$ accented word versus deaccented word) was manipulated to see the effect of accenting an item on recall of that item. Production data of the participants reading short passages were also collected to examine the link between the speakers’ perception and production. Speer and Foltz found that participants who used $L+H^*$ on focused items when reading aloud had faster response times to $L+H^*$ accented auditory stimuli than deaccented auditory stimuli. $H^*$ users, however, had no difference in response times between $L+H^*$ and deaccented stimuli. These results show that despite the fact that $L+H^*$ is canonically the only pitch accent type that is associated with focus, there is still variation between subjects as to whether they use a $L+H^*$ or $H^*$ pitch accent. This variation between subjects also appears to be linked to their perception of intonation.

In the processing literature, studies have also found that individuals differ in cognitive processing style. Personality traits that are associated with the Autism Spectrum Disorder, as measured by
the Autism Spectrum Quotient (AQ) questionnaire (Baron-Cohen et al., 2001), have been found to modulate the perception of prosodic cues in neurotypical individuals. The AQ questionnaire is used to assess the number of autistic traits an individual has. It is a self-administered survey containing a series of statements which are associated with one of five traits/domains (i.e., social skills, attention switching, attention to detail, communication, and imagination). People with a high AQ scores tend to have poor social skills, difficulty switching their attention, a keen focus on details, poor communication skills, and difficulty imagining situations or putting themselves in others’ shoes. The opposite is true for individuals with low AQ scores. For the survey, the participants read 50 statements and responded with how much they agreed with each statement (options are “Definitely agree”, “slightly agree”, “slightly disagree”, and “definitely disagree”). The responses of an individual are then used to calculate their overall AQ score, as well as scores for the five subscales which correspond to the different traits. The Communication subscale is of particular interest in this study because it has been show to relate to pragmatic ability, where a higher score means greater difficulty interpreting pragmatic meaning (Bishop, 2012). The Communication subscale statements include, “I enjoy social chit-chat” and “I find it easy to ‘read between the lines’ when someone is talking to me.” Disagreement with these statements would reflect a lower level of pragmatic skill, producing a low AQ score. In this paper, “pragmatic ability” is defined as the ability to ascertain the mental state or beliefs of a speaker given a context and what the speaker has just said.

It is important to note that these traits are being examined in people who are not diagnosed on the autism spectrum. Additionally, the AQ questionnaire is not a diagnostic tool. Despite this, the pragmatic difficulties we see in autistic population (Paul et al., 2005) are comparable with typically developed adults with high AQ scores (Baron-Cohen et al., 2001). The connection between autistic traits in typically developed adults and intonation seems plausible given what is known about the intonation of individuals diagnosed on the autism spectrum. Individuals with autism typically have difficulties with the perception and production of intonation, as well as pragmatic difficulties (McCann and Peppé, 2003), and since intonation is related to discourse meaning, this could explain
difficulties in producing and perceiving different prosodic contours that reflect different discourse meanings.

Jun and Bishop (2015a) found in their study that AQ score explained the variations found in the perception of prosody across individuals. Jun and Bishop used the structure priming paradigm to influence the prosodic structure readers project onto what they read, as a way of testing the Implicit Prosody Hypothesis. Fodor (2002) proposed that when reading silently, we project a default, language specific implicit prosody onto the text, and that implicit prosody guides the parser to favor the syntactic analysis associated with the default prosodic contour for the construction. This is known as the Implicit Prosody Hypothesis (IPH). The IPH has been used to explain cross-linguistic variation in attachment preferences in ambiguous sentences like (11). In languages, like Spanish, which prefer to attach high (i.e., the relative clause (RC) modifies the first noun (the servant) in the complex head noun), the default implicit prosodic contour would favor a high attachment reading. In languages like English which prefer to attach low (i.e., the RC modifies the second noun (the actress)), the default implicit prosodic contour would favor a low attachment reading. More specifically, an implicit prosodic break after the first NP in a sentence like (11) would result in attachment of the RC to the second NP (low attachment), while an implicit prosodic break after the second NP would result in attachment of the RC to the first NP (high attachment).

Jun and Bishop attempted to prime high attachment or low attachment of ambiguous relative clause sentences, like (11), using primes that were unambiguous restrictive clauses with no comma, (12), or unambiguous non-restrictive clauses with a comma, (13).

(11) The officer shot the servant$_{NP_1}$ of the actress$_{NP_2}$ who was on the balcony.

(12) The clumsy plumber changed the sink faucet that we installed.

(13) The clumsy plumber changed the sink faucet, which we installed.

The interesting finding was that the higher the Communication score on the Autism Quotient questionnaire (indicating a greater number of autistic-like communication traits), the stronger the priming effect that participants experienced. This was explained as higher AQ individuals possibly
not generating their own implicit prosodic contours for the target sentences, but rather high AQ participants were reusing the prosodic structure of the primes.

In another study of the priming of explicit prosody on relative clause (RC) attachment, Jun and Bishop (2015b) found that individuals who were high on the Communication AQ subscale were more likely to resolve attachment ambiguities with high attachment to the first NP (i.e., ‘the servant’ in (11)) than with low attachment to the second NP (i.e., ‘the actress’ in (11)) when they were primed with explicit prosody where a big prosodic break was after the second NP. Low AQ individuals, however, did the opposite and attached low when they heard a prosodic break after the second NP. The IPH predicts that if there is a large prosodic break after the second NP in sentences like (11), the the parser will attach the RC high to the first NP. That means that the low AQ individuals were doing something that was unexpected. Jun and Bishop proposed that low AQ individuals are probably more sensitive to the phonological prominence of the NP before the prosodic break. That is, since the second word is the last accented word in the intermediate phrase (i.e., the Nuclear Pitch Accented (NPA) word), which is phonologically the most prominent word, the low AQ individuals must have attached the RC to the prominent NPA word in that phrase.

2.5 Hypotheses and Predictions

Pragmatic ability, as measured by the AQ questionnaire, and production seem to have some bearing on perception in a systematic way. According to the results of Speer and Foltz’s study, L+H* users were faster to response to words that carried a L+H* than words that carried no accent. On the other hand, H* users showed no difference in response times between the accented and deaccented conditions. For individuals with high AQ scores, they appear to be more sensitive to prominence than individuals with low AQ scores (Bishop, 2013a; Jun and Bishop, 2015a,b). In order to study the differences in perception of H* versus L+H* and their integration in processing of meaning, this paper will adopt Ito and Speer’s design, whereby L+H* and H* are predicted to have distinct looking patterns. We examine the relationship between production (i.e. speaker’s choice of pitch accent type when marking focus) on perception of L+H*, as well as the effect of AQ
The variation in perception presents itself in the extent to which participants exhibit facilitatory effects of felicitous L+H* use in a visual search task or garden path effects with infelicitous L+H* use, which Ito and Speer found.

It is predicted based on previous findings that L+H* users should be more likely to be sensitive to the L+H* than Non-L+H* users. Therefore, L+H* users should be fastest to find the target in felicitous L+H* conditions (i.e., “Find the blue ball. Now, find the GREEN_{L+H*} ball.”) and slowest, because of being garden pathed, when hearing an infelicitous L+H* (i.e., “Find the blue chair. Now, find the GREEN_{L+H*} ball.”). The H* users would not attend to the prominence difference between H* and L+H*, and therefore would have a smaller difference in the speed to find the target object between the two conditions. They are not expected to make predictive, anticipatory looks before the onset of the noun because they are not attending to the intonation, or they are less likely to attend to it.

Regarding AQ, low AQ individuals should be more likely to show differences between L+H* and H* than high AQ individuals. Because low AQ individuals have been found to be more sensitive to prominence, they should have no difficulty integrating the prosodic cues to focus into their processing of the stimuli. High AQ individuals, on the other hand, should have difficulties decoding the use of L+H* or H* into its associated pragmatic meaning, and therefore have similar speeds of finding the object in felicitous L+H* and infelicitous L+H* conditions.

First, in order to determine the participant’s ability to encode pragmatic meaning using intonation, a production experiment was conducted. It tested participants’ sensitivity to the information status of a word and the type of pitch accents employed to mark it as focused. Participants had to produce adjective-noun pairs that were in contrastive focus and pitch accent types were labeled to determine how they marked contrastive focus phonologically. This study limits the scope of encoding and decoding of pragmatic meaning to only contrastive focus.
3. Experiment 1: Production

3.1 Participant information

128 UCLA students participated in this study. Participants were given course credit or monetary compensation for participating in the study. Only data from 105 participants were subjected to analysis. The data of 23 participants could not be used because of either experimenter error (n=13), not being a native English speaker (n=9), or loss of hearing in one ear (n=1). In line with previous linguistic work which investigated individual variations based on autistic-like traits in neurotypical individuals (Bishop, 2012, 2013b; Yu et al., 2013; Jun and Bishop, 2015a), the autism spectrum quotient (AQ) questionnaire (Baron-Cohen et al., 2001) was used. Participants completed the questionnaire after both experiments were completed so as to keep the experimenters blind to the participant’s AQ score and minimize experimenter bias. Only the Communication subscale is reported because this was the trait of chief importance in previous studies and it is the trait most closely associated with pragmatic ability, which is needed in encoding and decoding intonational meaning. The distribution of Communication AQ scores of participants is shown in Figure 2. Although participants’ AQ scores were treated as a continuous variable, for easy of visualization they were grouped into high, mid, and low AQ groups. The high AQ group corresponds to participants with an AQ score greater than 1 SD above the mean (i.e., a score of 23 or greater) and the low AQ group corresponds to participants with an AQ score greater than 1 SD below the mean (i.e., a score of 15 or less).

3.2 Materials

Images for the study were taken from the Bank of Standardized Stimuli (BOSS) database (Brodeur et al., 2014). This database contains 1468 photos of various objects (animals, toys, fur-
Figure 2: The distribution of Autism Quotient scores of participants (\( \bar{x} = 19, \sigma = 4.26 \)). AQ scores greater than 22 (approx >+1 SD above the mean) were considered high AQ, and scores lower than 16 (approx <-1 SD below the mean) were considered low AQ.

The distribution of Autism Quotient scores of participants (\( \bar{x} = 19, \sigma = 4.26 \)). AQ scores greater than 22 (approx >+1 SD above the mean) were considered high AQ, and scores lower than 16 (approx <-1 SD below the mean) were considered low AQ.

3.3 Procedure

Before the experiment began, participants were familiarized with all of the names of all of the objects in each of the four colors. The production task was administered first so that the participants’ own productions would not be influenced by the auditory stimuli they heard from the eye tracking task.

After consenting and familiarization of the images, participants were seated in a sound attenuated booth in front of a computer monitor. Each trial consisted of two parts. In the first part, the

---

1The elephant is the exception to this group. It was chosen because it was easily identifiable and because of its sonorant consonants, which result in more reliable pitch tracking
participant was presented with one of 8 objects (i.e., ball, bow tie, chair, crayon, elephant, flower, funnel, pillow) in one of 4 colors (i.e., blue, green, orange, yellow). The participant then named the image (see Figure 3) they saw using an adjective-noun sequence in the carrier phrase, “There is ______.” (e.g., There’s a green pillow.). Participants were free to use the indefinite article, “a” or “an”, when naming the image. After naming it, they would move on to the second part of the trial. In this part, the previous image would be removed and replaced with another image (see Figure 4). Again, the participant named the image they saw using an adjective-noun sequence, but they also named the previous image (e.g., There was a green pillow, but not there’s a blue ball). This was done to highlight the contrast between the previous and current images. Participants produced the color and type of object in the image in the carrier phrase, “There was a _____ but now there’s a ______.”

All trials were presented in a randomized order. There were 24 trials in total with 16 filler trials. The filler trials consisted of naming two images that appeared concurrently.

Figure 3: Block 1 non-contrastive trial. What was seen in the first half of the trial is on the left, and the second half of the trial appears on the right.
Figure 4: Block 1 contrastive trial. What was seen in the first half of the trial is on the left, and the second half of the trial appears on the right.

3.4 Results

Data from 8 participants could not be used due to experimenter error, leaving 8 trial halves (The first half of the trial and the second half of the trial were cut into separate recordings) for each of 97 participants, resulting in 388 recordings contributing to the analysis.

The tunes elicited for contrastive focus varied not only phonologically, but also phonetically. For example, 31 different combinations of pitch accents (i.e., tunes) were used on the adjective-noun pair that was in focus. Figure 5 shows the frequency of each pitch accent type used, in a discourse context where the adjective is focused, on the adjective (left) and noun (right). For adjectives, the most common type of pitch accent is the L+H*. It was used 40% of the time in the focus condition, which is a similar rate to what has previously been found (Ito and Speer, 2006). The next most used accents were H*, H+H* and L*. The choice of the pitch accent type seems to reflect relative prominence of the pitch accents. In her dissertation, Ayers (1996) also found a hierarchy of prominence of NPA words with an expanded pitch range (L+H*), NPA words that were normally accented (H*), and deaccented words. The stop initial words that were accented

---

2 Any accents or tunes that were used less than 5% of the time were not included in the graphs, but were counted in calculation of percentages.

3 The !H* can sometimes be phonetically very similar to the L*, especially after a succession of downstepping or by virtue of being phrase final. Because of these cases we consider L* and !H* to both be equally low in prominence.
in one of those 3 respective categories had descending values of VOT on the initial stop. Ayers took this to mean that the L+H* on the NPA word was the most prominent pitch accent of all, followed by the H* NPA word, and then the deaccented word. From an information structure standpoint, we would expect the L+H* to be used most often to narrowly focus an object (e.g. a word in contrastive focus) because the nature of focus is to make an item salient, relative to some alternative or set of alternatives.

Figure 5: The distribution of pitch accents used by participants on the adjective (left) and on the noun (right) respectively in order to signal contrastive focus on the adjective. The ‘*?’ label is meant to indicate uncertainty about whether a word is accented or not.

The prominence relation was the opposite for the preferred accent on the noun. Around 60% of the time, participants chose to deaccent the noun. The other pitch accents used for the post-focus noun ascend in prominence as they descend in use. This deaccenting of the words after a focused word makes that focused word the NPA word. This makes the adjective the phonologically most prominent word in the phrase (Ladd, 2008). Therefore, not only does deaccenting serve to indicate that the noun is given information, but it also enhances the prominence of the pitch accent on the adjective.
Figure 6: The various tunes employed over the adjective-noun pair in order to mark focus on the adjective.

Figure 6 shows the frequency of the sequence of accents (or tunes) on the adjective and noun. The tune that is most often used (30% of the 388 productions) was a \(L+H^*\) on the adjective, followed by a deaccented noun (henceforth to be represented by "0"). This tune is the canonical tune for focus in English (Pierrehumbert and Hirschberg, 1990). The second (\(L+!H^*\) 0; 10%) and third (\(H^*\) 0; 8%) most used tunes also exhibit the deaccenting of the noun after the focused adjective. As mentioned before, this deaccenting of the noun makes the adjective the NPA word. The \(L+H^*\), which is already the most prominent accent, becomes even more salient by carrying the nuclear pitch accent. With regard to adjective pitch accent choice of the three most used tunes, they are descending in prominence as well (i.e., \(L+H^*\), \(L+!H^*\), \(H^*\), etc.).

Speakers attempted to mark focus by increasing the relative prominence difference between the adjective and the noun by employing very prominent pitch accents on the adjective and weakly prominent pitch accents on the noun (or often deaccenting the noun entirely).

With regard to the AQ Communication subscale and production, there seems to be little variation between AQ groups. Table 1 and Figure 7 show the proportions of use of the different pitch accent types (and tunes) on the adjective, the noun, and adjective-noun sequences by AQ score group. All pitch accent types and tunes were used by participants in the 3 groups. The relative proportion of use of pitch accents is also similar across AQ scores. In the left table in 1, the proportion of utterances that had a particular accent type on the adjective is shown, while on the right table,
The distribution of the proportion of pitch accents used, for various AQ scores, on the adjective in order to signal contrastive focus on the adjective.

Table 1: The proportion of utterances that had a particular accent type on the noun is shown. With regard to the types of pitch accents used on the adjective, the low AQ group used L+!H* about 10% more than the mid and high AQ groups. However, if you add up the proportion of L+H* and L+!H* uses in the three group, they all add up to be about the same (around 55%). For the noun that appears after focus, low AQ individuals tended to deaccent the noun more often than mid and high AQ individuals. The mid and high AQ groups compensated for this decreased use of deaccenting by using the !H* accent on the noun often often.

The bars in Figure 7 do not reach 100% because only the tunes that were used at least 5% of the time were plotted. Based on the bar heights, high AQ individuals seemed to more often use the lesser used tunes, followed by mid AQ individuals, and the low AQ individuals.

This variation in the types of pitch accents used might be due to noise from task difficulty. Participants reported trouble naming certain colors and types of objects despite the fact that they were familiarized with all of the images used before beginning the experiment. Not only were
some types of objects named incorrectly, but there were a number of disfluencies as well. It could be the case that the participants tried reading the carrier phrase each time, instead of thinking about the image and naming it naturally with the phrase “There’s a _____.” If participants were simply reading the carrier phrase out loud, their intonation might have been affected by a desire to read the text fluently, regardless of pragmatic meaning, which has been found to be a problem in production studies where production prompts presented as written text (Jun, 2010). Images of the objects to be named replaced their corresponding text in the carrier phrase as a way to discourage participants from strictly reading text. Images were also included in the carrier phrase text to ease memory load and also highlight the intended alternative.

Because of the variation within and across speakers with regard to their choice of pitch accent
type for focus, participants could not be easily categorized as H* users or L+H* users for the purposes of modeling the results in Experiment 2. One alternative to categorizing participants’ productions in this fashion would have been to have one group for each type of pitch accent used (i.e., H* users, H+!H* users, L* users, etc.). Many of these groups, however, would have little data in them, and so many groups would greatly reduce the power of the model. Additionally, the frequency of L+H* use could be used. The issue of that, however, is that with such few trials (4), it becomes hard to interpret what kind of differences there are between participants who use a L+H* once versus those that used it twice. Perhaps with more trials pitch accent type usage could be treated numerically. Because of the inappropriateness of the other methods of categorization, participants were grouped by whether they used a L+H* at least once (L+H* users) or they never used a L+H*, (Non-L+H* users).
4. Experiment 2: Perception

In the second experiment, participants engaged in a visual search task which tested their sensitivity to the differences between H* and L+H* in focus and discourse new contexts. After being presented with an array of images of different types of objects of different colors, they heard an instruction to find one of those images. After finding and clicking on it, a second instruction would play, telling them to find another image. By manipulating the discourse context created by the instruction sequences (i.e., a sequence of the same noun twice or two different nouns) and manipulating the type of pitch accent that was on the adjective, we would be able to see the extent to which participants interpreted L+H* as indicative of contrastive focus, and thereby prompting anticipatory eye movements to the upcoming noun. If the prosodic cues were being attended to and integrated into processing, anticipatory looks to the previously named type of object cell would occur when there is a L+H* on the adjective. On the other hand, a H* on the adjective should not lead to anticipatory looks to the previously mentioned object cell. This would be a direct replication of Ito and Speer’s findings.

4.1 Visual and Auditory stimuli

The same color and object images that were used in Experiment 1 were used in Experiment 2. Each adjective noun pair was recorded in the carrier phrase, “Find the ADJECTIVE NOUN” by a trained phonetician. Hence forth, phrases like “Find the blue ball” will be referred to as “instructions.” Once the instructions were recorded, the best production of the adjective and noun for each condition were spliced together, forming adjective-noun 64 pairs (2 pitch accent types (H* vs. L+H*) × 4 colors × 8 objects). In addition to this, another set of recordings were made to serve as the first set of instructions to be heard in a trial. The first set of instructions set up a contrastive discourse context or not depending on whether the noun was repeated in the second
After splicing the recordings together, the f0 was manipulated for the L+H* and H* conditions in Praat. For the L+H* condition (i.e., L+H* on the adjective, and no accent on the noun), the f0 started in the middle of the speaker’s pitch range (145Hz), fell to a low target (110Hz) in the article “the”, and then rose to a high peak (250Hz) at the end of the stressed syllable of the adjective (see Figure 8). The noun following the L+H* adjective was always deaccented with very low f0 and creaky voice, so its f0 was not manipulated. For the H* condition (H* on the adjective and !H* on the noun), the f0 started at the middle of the speaker’s pitch range (145Hz) and rose to a high peak (195Hz) at the end of the stressed syllable of the adjective (see Figure 9). The f0 then fell to 160Hz near the beginning of the stressed syllable of the noun. The intensity of all stimuli were then normalized using a Praat script and durational differences were controlled for between pitch accents. The contour of the first instruction was always a H* on “find”, followed by a H* on the adjective, and a !H* on the noun (see Figure 10).

Figure 8: Pitch contour and spectrogram of the instructions for the L+H* condition.
Figure 9: Pitch contour and spectrogram of the instructions for the H* condition.

Figure 10: Pitch contour and spectrogram of the first instruction set that created the discourse context.
4.2 Eye tracking task

Participants seated in front of an SR Eyelink 1000 eye tracker and monitor. After the calibration routine, the participant was explained the task they would have to do and given a pair of noise canceling headphones. The task was to find and click on an object that was named through the headphones. On the monitor, the participants saw an array with four object cells (not all objects were in every visual scene). Before the onset of the first instruction, the participant was given a 2000ms preview of the array so that they could have ample time to become familiar with all of the types of objects, their colors, and their locations, as they all varied across trials. An example array can be found in Figure (11). After the preview, the first auditory stimuli file (i.e., the context phrase) was played instructing the participant to find a particular image in the array.

After finding and clicking on the image, there was a 3000ms delay before the onset of the discourse marker "now". This was followed by the second instruction. The critical manipulations on the accent of the adjective were made in the second instruction. The discourse context varied based on the noun in the first instruction and the pitch accent varied on the adjective in the second instruction (2 contexts × 2 pitch accent types). This resulted in four conditions, as seen in (14-17) (capitalized words indicate a L+H* pitch accent).

When the nouns in the two instructions were the same but the adjectives differed, this put the adjective in the second instruction in focus. This focus context allows for the felicitous use of L+H*, (16), and the infelicitous use of H*, (15) because H* marks new information. When the nouns and the adjectives both differed, this favored the use of the H* to mark the adjective and the noun as new information. Therefore, that context was felicitous for H*, (14), and infelicitous for L+H* (17).
(14) **Felicitous H*\**

1st: “Find the blue chair.”
2nd: “Now, find the green\textsubscript{H*} ball\textsubscript{H*}”

(15) **Infelicitous H*\**

1st: “Find the blue ball.”
2nd: “Now, find the green\textsubscript{H*} ball\textsubscript{H*}”

(16) **Felicitous L+H*\**

1st: “Find the blue ball.”
2nd: “Now, find the GREEN\textsubscript{L+H*} ball\textsubscript{H*}”

(17) **Infelicitous L+H*\**

1st: “Find the blue chair.”
2nd: “Now, find the GREEN\textsubscript{L+H*} ball\textsubscript{H*}”

Again, the participant found and clicked on the target image. After a 500ms delay from the click, the experiment would advance to the next trial.

A Latin square design was used to ensure that no participant saw the same item in multiple conditions. Each participant experienced 36 total trials, with two-thirds of those being “felicitous” trials (i.e., 12 felicitous H*, 12 felicitous L+H*, 6 infelicitous H*, 6 infelicitous L+H* trials). This was done to avoid participants finding the intonation unhelpful, and thus attending less to it. If too many infelicitous trials occur across the experiment, participants may adopt strategies to inhibit their attention to prosodic cues.
Figure 11: An example of an array of objects used in the eye tracking task.
4.3 Results

The eye fixation data was recorded at 500 Hz (sampled every 2ms) and then later binned into 50ms time bins. If a fixation occurred in a given area of interest (i.e., “ball” cell), then the fixation was coded as 1. If a fixation did not occur in that given area of interest in that time bin, it was coded as 0. Out of the total 3780 trials (105 participants × 36 trials) collected, 347 trials (9.2%) were excluded where the participant fixated on the target cell before the onset of the adjective and remained there into the adjective. It takes about 200 milliseconds to plan and execute a saccade (Rayner, 1998), and therefore any fixations to the target in that time window are assumed to be random fixations. Since the participant was already fixating in the target cell before the onset of the adjective and remained there, this tells us nothing about whether they were anticipating the upcoming noun based on the information in the adjective or not.

The time window of interest was 200ms after the onset of the adjective until 500ms after the offset of the noun. 500ms after the offset of the noun was chosen as the cutoff by examining the grand mean for when participants hit ceiling with regard fixating on the target noun cell. Barr (2008) recommends this method of analysis window selection so as not be unduly influenced by the looking patterns of individual conditions when selecting the analysis window.

The lme4 package was used to run a mixed effects logistic regression using the glmer function (Bates et al., 2015). For the purposes of the model, the production data of participants was used to group them by their usage of L+H*. One group represented L+H* users, who used a L+H* or L+!H* at least once out of the four trials, and non-L+H* users, who never used a L+H* in any of the four trials. Categorizing speakers as L+H* users or non-L+H* users was chosen because there is a greater distinction to be made between those two groups than two groups where participants either always used a L+H*, or did not always use a L+H*. The latter type of grouping would result in a group shared by those who use a L+H* three out of four times, and those who never used it. Speaker who use L+H* at least once (i.e. speakers who use L+H* 4/4, 3/4, 2/4, and 1/4 times) have more in common in terms of pitch accent choice than speakers who don’t always use L+H*.
In the first round of modeling, TIME, CONDITION, pitch accent use and AQ score were all entered as fixed effects, including all interactions. In line with Barr et al. (2013), maximal slopes and intercepts of TIME, CONDITION, and TIME × CONDITION for participant and item were included as random effects. This maximal model did not converge however, so a model with maximal intercept and slope of TIME and CONDITION, but not their interaction, was used. The highest order fixed effects were removed one by one and compared to the larger model with that fixed effect included. Only fixed effects that significantly improved model fit, relative to a subset model, were included. The resulting model did not include the Communication AQ subscale score (see Table 2 for model comparisons) or pitch accent use at all because their inclusion did not significantly improve model fit. Table 3 show the results from this model which does not include AQ or pitch accent use. The levels of CONDITION were dummy coded with the reference level being felicitous H*1.

Figure 12 shows that the odds of looking to the target cell at the earliest time point, 200ms into the adjective, are very low. The log odds are around -2, which when exponentiated gives the odds of 0.135. That means that the odds of looking to the target were about 1:7.

Main effects of TIME and CONDITION There was a significant main effect of TIME, meaning that as time went on in the FELICITOUS H* condition, the log odds on looking to the target increased ($\beta = 4.459, p = <.001$). Compared to FELICITOUS H*, the log odds of looking to the target were overall higher in the FELICITOUS L+H* condition ($\beta = 1.895, p = <.001$). This is to be expected since the felicitous use of L+H* is predicted to allow the processor to predict the upcoming noun according to Ito and Speer’s results. This results in more looks to the target cell sooner than in the felicitous H* case. The log odds of looks to the target in the INFELICITOUS H* condition ($\beta = 1.664, p = <.001$) were higher than in the FELICITOUS H* condition. The log odds of looking to

1 The following glmer formula was used for the model:

```r
fixation ~ time * condition + (1 + condition + time|participant) + (1 + condition + time|item)
```
<table>
<thead>
<tr>
<th>Model Predictors</th>
<th>Difference in df</th>
<th>$\chi^2$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>TIME + COND + AQ + TIME:COND + TIME:AQ + TIME:COND:AQ</td>
<td>3</td>
<td>1.78</td>
<td>0.6195</td>
</tr>
<tr>
<td>vs. No TIME:COND:AQ</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TIME + COND + AQ + TIME:COND + TIME:AQ</td>
<td>1</td>
<td>0.46</td>
<td>0.4992</td>
</tr>
<tr>
<td>vs. No TIME:AQ</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TIME + COND + AQ + TIME:COND</td>
<td>1</td>
<td>0.35</td>
<td>0.5537</td>
</tr>
<tr>
<td>vs. No AQ</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2: Comparisons of model fits for Autism Spectrum Quotient score as a predictor.

<table>
<thead>
<tr>
<th>Factor</th>
<th>$\beta$</th>
<th>SE</th>
<th>z</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-5.778</td>
<td>0.152</td>
<td>-37.89</td>
<td>&lt; .001 *</td>
</tr>
<tr>
<td>TIME</td>
<td>4.459</td>
<td>0.118</td>
<td>37.88</td>
<td>&lt; .001 *</td>
</tr>
<tr>
<td>COND (FELICITOUS L+H*)</td>
<td>1.895</td>
<td>0.137</td>
<td>13.85</td>
<td>&lt; .001 *</td>
</tr>
<tr>
<td>COND (INFELICITOUS H*)</td>
<td>1.664</td>
<td>0.154</td>
<td>10.80</td>
<td>&lt; .001 *</td>
</tr>
<tr>
<td>COND (INFELICITOUS L+H*)</td>
<td>-0.006</td>
<td>0.166</td>
<td>-0.40</td>
<td>0.971</td>
</tr>
<tr>
<td>TIME $\times$ COND (FELICITOUS L+H*)</td>
<td>-0.864</td>
<td>0.084</td>
<td>-7.06</td>
<td>&lt; .001 *</td>
</tr>
<tr>
<td>TIME $\times$ COND (INFELICITOUS H*)</td>
<td>-0.805</td>
<td>0.102</td>
<td>-5.77</td>
<td>&lt; .001 *</td>
</tr>
<tr>
<td>TIME $\times$ COND (INFELICITOUS L+H*)</td>
<td>-0.013</td>
<td>0.110</td>
<td>-0.12</td>
<td>0.908</td>
</tr>
</tbody>
</table>

Reference level for CONDITION = FELICITOUS H*

Table 3: Estimates, standard errors, z values, and p values of best fitting model
Figure 12: Log odds of looking to the target object cell for each condition.

The target cell in the INFELICITOUS L+H* condition were nearly identical ($\beta = -0.006, p = 0.971$) to FELICITOUS H*.
Interaction of **TIME × CONDITION**  The interaction between **TIME** and **CONDITION** relates to the difference in slope between conditions. This means that near-zero coefficients indicate that the slopes are the same (parallel lines), non-zero positive coefficients mean the difference in the log odds of target fixations in the conditions is getting larger over time, and a non-zero negative slope means the difference in log odds of target fixations is getting smaller over time. The model confirms what is evident from just looking at Figure 12; FELICITOUS H* and INFELICITOUS L+H* are changing at the same rate ($\beta = -0.013$). Also, the slope for FELICITOUS H* is higher than the slope for FELICITOUS L+H* ($\beta = -0.864, p = <.001$) and INFELICITOUS L+H* ($\beta = -0.805, p = <.001$). What this means is that the differences in looks to the target cell between the felicitous H* condition and the felicitous L+H* and infelicitous H* conditions reduce over time.

**Anticipatory looks**  Replicating Ito and Speer’s results, the looks to the target object cell in the FELICITOUS L+H* condition started to increase earlier than looks to target in the FELICITOUS H* condition. The FELICITOUS L+H* condition elicited an increase in looks that started before the onset of the noun. What this indicates is that the processor is predicting the upcoming noun referent based solely on the prosodic cues, before any segmental information is available, and planning eye movements in anticipations. On the other hand, the FELICITOUS H* condition was no different from the INFELICITOUS L+H* condition. This is at odds with Ito and Speer’s results which suggested that the increase in looks to the target in the INFELICITOUS L+H* would start later than in the FELICITOUS H* condition. This is because the infelicitous use of the L+H* provides misleading prosodic information which causes the processor to incorrectly predicate the upcoming noun, which means it then has to change the gaze location once the noun’s segmental information is available. This takes time to do, resulting in a lag in reaching the target. This lag should not happen in the H* condition because the H* should give the processor no additional information as to the identity of the noun referent. Therefore, only the effect of felicitous L+H* being faster than infelicitous L+H* is replicated in this study.
AQ and Pitch Accent use  As mentioned before, the addition of AQ and pitch accent use did not significantly improve model fit, so they were left out of the model. A visual inspection of the looking patterns for these two factors shows that there is no clear relationship between the AQ (see Figure 13) or pitch accent type use (see Figure 14) and fixations to the target. The logs odds of looking to the target are nearly identical for high, mid, and low AQ groups in all four conditions. The only AQ group that seems to differ from the other two is the high AQ group in the INFELICITOUS L+H* condition. In that condition, the high AQ group seems to have non-significantly higher log odds of fixations to the target than the mid and low AQ groups. Figure 14 does not suggest a strong link between perception and production of intonation either. Both groups have very similar looking patterns in all conditions.

![Log odds of fixations to Target cell](image)

Figure 13: Log odds of looking to the target object cell for each condition for high, mid, and low AQ groups.
Figure 14: Log odds of looking to the target object cell for each condition for L+H* users and Non-L+H* users.
4.4 Garden Pathing effect

In Figure 15, there seem to be an effect of the infelicitous use of L+H* in the right panel. Before the onset of the noun, participants seem to be looking to the competitor object rather than the target. It must be the case that they are predicting that the upcoming noun will be the same in the previous instruction, and then changing their gaze once they realize that they have made an erroneous guess. This garden path effect accounts for the delayed looks to the target cell, since participants are first looking to the incorrect competitor cell. Once the participant processed enough of the segmental information of the noun, the looks sharply dropped off, and looks to the target object cell increased. That same garden path effect, however, occurred in the felicitous H* condition (see the “fel H*” condition in the left panel in Figure 15). Here, log odds of looks to the competitor increase until around 150ms into the noun. This earlier trade off between target and competitor looks suggests that participants needed more segmental information from the noun before deciding that they had incorrectly predicted the upcoming object when hearing a L+H* compared to when they heard the H*. This could be because they committed to their prediction more strongly with the stronger acoustic cues to focus (i.e., L+H* as opposed to H*), or simply that there was variation among participants as to whether they kept entertaining the competitor as an option. While this garden path effect is expected for the infelicitous L+H* condition, it is unexpected for the felicitous H* condition, which was supposed to be a baseline condition. The results show that the looking pattern to the target object cell in the felicitous H* condition is very similar to the infelicitous L+H* condition.

Figure 16 shows the log odds of looking to the target object cell, as well as the competitor object cell, in the felicitous L+H* (in the left panel) and infelicitous H* (in the right panel) conditions. These looking patterns are nearly identical, and show that at no point are participants more likely to consider the competitor than the target. The only difference comes out in the looks to the competitor cell after the noun offset. It seems that looks to the competitor object reach a log odds of -6, or 1:402 odds of looking to the target, about 300ms after the offset of the noun in
Figure 15: Log odds of looking to the Target and Competitor object cells for Felicitous H* and Infelicitous L+H* conditions.

Figure 16: Log odds of looking to the Target and Competitor object cells for Felicitous L+H* and Infelicitous H* conditions.
the felicitous L+H*, but the log odds do not drop as low in the infelicitous H* condition. This could indicate participants lack of confidence in ruling out the alternative object of the same color, whereas the alternative is clearly rejected quickly in the felicitous L+H* condition.

4.5 Post-Hoc Experiments

Two post-hoc experiments were conducted to address the issue of possible f0 abnormalities in the H* stimuli. Since participants seemed to treat the H* stimuli the same as the L+H* stimuli, it is possible that the H* stimuli were too similar to the L+H* stimuli in terms of f0. Therefore, two new versions of the eye tracking experiment were run which tested different versions of the H* stimuli, which were resynthesized. In the first experiment, the H* stimuli were manipulated such that the 40Hz rise in the original stimuli was reduced to a 10Hz rise by raising the initial f0 15Hz, and lowering the peak f0 by 15Hz. This version of the experiment will be referred to as the “Small Rise” version. The second version used the falling pitch accent, H+!H*, on the adjective instead of a H*2. This pitch accent type was chosen because of its phonetic dissimilarity to L+H*. This version of the experiment will be referred to as the “Falling Accent” version. In both versions, the L+H* stimuli remained the same.

4.5.1 Methods and Procedure

18 UCLA students participated in this study. Participants were given course credit or monetary compensation for participating in the study. The data of 3 participants had to be excluded from analysis because of experimenter error, resulting in usable data of 15 participants. Participants first did the “Small Rise” version of the experiment followed by the “Falling Accent” version. The procedure was the same as Experiment 2 except that participants did not complete the AQ questionnaire.

2Technically, the H* and H+!H* are not interchangeable. The discourse context would have to change to make the use of H+!H* truly felicitous. For the purposes of the post-hoc experiment however, the use of the H+!H* the strictly motivated by its phonetic dissimilarity to the rising L+H* pitch accent. We were trying to elicit a different looking patterns with a phonetically different accent since the original H* and L+H* were both rises.
4.5.2 Results and Discussion

Just as in Experiment 2, there did not seem to be a difference between pitch accents types. In the “Small Rise” version (see Figure 17 for results), looks to the target object cell were the same for the felicitous small rise H* condition and infelicitous L+H* condition. Also, the felicitous L+H* condition appeared to be the same as the infelicitous small rise H* condition. In the “Falling Accent” version (see Figure 18 for results), there was slightly more deviation. In the infelicitous falling accent H* (H+!H*) conditions, participants didn’t seem to be quite as fast at looking to the target as in the felicitous L+H* condition. The felicitous falling accent H* and the infelicitous L+H* accent were also the same. This suggests that the participants were not relying on the f0 cues as their primary cue for prediction. Expectations about the experiment might have influence their behavior and obscured an effect of pitch accent type.

![Figure 17: Log odds of looking to the target object cell for each condition.](image-url)
Figure 18: Log odds of looking to the target object cell for each condition.
5. Discussion and Conclusion

Although this study was modeled after the design of Ito and Speer’s visual search task, their results were only partially replicated. There were indeed anticipatory looks to the target object cell in the felicitous L+H* condition, and a garden pathing effect in the infelicitous L+H* condition, but the conditions with H* stimuli patterned very differently from previous findings.

One possibility for the failure of this replication is that participants paid no attention to the intonation and instead always thought that the upcoming object would be the same noun as in the first instruction. This is supported by the fact that we see a rise in looks to the previously named object cell in every condition (see the red target line in Figure 16 and the blue competitor line in Figure 15). To address this, two post-hoc experiments were done that manipulated the f0 of the original H* stimuli in an effort to elicit different patterns of looks than the L+H* stimuli. Both version of the H* stimuli failed to do so. This suggests that participants were not attending to the intonation in any meaningful way.

In debriefing, some participants reported a strategy in which they would keep their eyes in the previously mentioned object cell because occasionally the same type of object would be named. It is the case that participants had a baseline preference to look at the previously mentioned object cell before the onset of the instruction. This means that participants kept looking to the previously object cell because they were expecting to be told to find different colored object of the same type. The results presented in this paper, however, excluded trials in which the participant was already looking at the target object in the first 150ms of the adjective. What this means is that after hearing 200ms of the adjective, participants then started to look to the previously mentioned object cell. This could be consistent with participants using the intonation to predict that the object would be the same, but it is also consistent with an account where participants expected the object to be the same, and were only waiting for the segmental information of the adjective to start looking to the
same object type as before, but with the color just named. This non-intonational account of their behavior explains why the realization of the "H*" in Experiment 2 and the post-hoc experiments did not matter.

Participants could have been ignoring the f0 differences between the H* and L+H* stimuli because of they sounded unnatural. The H* stimuli created may have sounded unnatural or ambiguous compared to the L+H* because the manipulation was done only in f0 but not in other acoustic properties. This would also help to explain the post-hoc results as well. Other acoustic cues to the difference between the accent types could have interfered with the processing of the stimuli. Breen et al. (2010) found that focused words are produced with higher intensity, high mean f0, higher peak f0, and greater duration than unfocused words. Because duration and intensity were controlled across pitch accent types in this study, those cues were unhelpful in distinguishing between L+H* and H*, and could have resulted in confusion on the part of the participant.

5.1 Individual Differences

While there were marginal differences in the pitch accent productions of high, mid, and low AQ groups, their looking patterns did not seem to differ. Including AQ as a predictor in the model did not significantly improve model fit, so it was not included in the final model. While autistic-like traits seems to influence the final parsing of a sentence through implicit prosody in Jun and Bishop’s work, it does not seems to effect the on-line processing of explicit prosody in this study. Perhaps if participants were asked about their interpretation of the second set of instructions (i.e., asking some question that probes at their interpretation of focus, after having heard the entire instruction), there would be an effect of AQ. None of the other studies related to AQ look at the time course of processing. Instead, they are focused on post-stimulus questions. This lack of an effect could also be due to the issues with the stimuli that were mentioned previously.

The results of this study did not find a link between perception and production as Speer and Foltz did. The L+H* users and Non-L+H* users both had similar looking patterns in all four conditions.
To conclude, the link between production and perception, and the effect of AQ were not found in this study. This could be for a number of reasons which has been previously discussed above. In order to better understand the results of this study, a follow-up would need to be conducted to better understand the acoustic features of H* and L+H*, other than f0, that lead participants to treat them as the same or different. With what can be said about the L+H* stimuli, it seems that participants of all AQ scores interpreted the L+H* as focus and used that for prediction. Therefore, it could be the case that AQ differences do not cause disruption of the integration of pitch information. High AQ individuals were just as likely to use L+H* for prediction as low AQ individuals. It could be that only the final interpretation is what is affected by difficulties in understanding pragmatics. Additionally, speakers who never use a L+H* do not seem to differ from speakers who do use L+H* when it comes to perception of focus. Further study is needed to understand the types of domains in which these AQ effects, as well as effects of the link between perception and production, appear and influence language production and comprehension.
Bibliography


