THE PRODUCTION AND PERCEPTION OF PROSODIC PROMINENCE IN URBAN NAJDI ARABIC

by

Hussain Almalki
A Dissertation
Submitted to the
Graduate Faculty
of
George Mason University
in Partial Fulfillment of
The Requirements for the Degree
of
Doctor of Philosophy
Linguistics

Committee:

_____ Director

_____

_____

_____ Department Chairperson

_____ Program Director

_____ Dean, College of Humanities and Social Sciences

Date: _____ Spring Semester 2020
George Mason University
Fairfax, VA

The Production and Perception of Prosodic Prominence in Urban Najdi Arabic

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at George Mason University

By

Hussain Almalki
Master of Arts
Florida International University, 2014
Bachelor of Arts
Al-Imam Mohammad Ibn Saud Islamic University, 2010

Director: Steven H. Weinberger
Department of Linguistics

Spring Semester 2020
George Mason University
Fairfax,

Dedication

This dissertation is dedicated to my wife, Dr. Sahar Almohareb, and our two children Meshal and Yara.

Acknowledgements

This work would not have been possible without the help, support, guidance, and patience of so many people. First, I would like to express my deepest and undying gratitude to my advisor, committee chair, and friend Dr. Steven Weinberger for his friendship, knowledge, and support. Thank you, Steven, for allowing me to continue working on this project and providing me with invaluable feedback and guidance throughout its stages. I want to express my heartfelt gratitude to my committee members Dr. Douglas Wulf and Dr. Harim Kwon, for their input and invaluable suggestions and comments. I would also like to express my appreciation to Dr. Cynthia Lukyanenko, Dr. Laurel Brehm, and Dr. Zhyian Gao for their help during the statistical analysis stage of this project.

Many thanks go to my friends: Abdulrahman Binhadlq, Abdulhadi Alfawzan, Abdullah Almushri, Faisal Albarrak, Mohammad Azzam Mustafa, and Meshal Alwehaibi for their friendship and support. I especially thank, Meshal Alwehaibi, for stepping in just when my family needed him. We truly appreciate you, Meshal, and we will forever be indebted to you. You are a God sent!

I want to express my deepest and most sincere gratitude and appreciation to my wife, Dr. Sahar Almohareb, for her resilience, patience, love, and kindness. Special thanks go to my son, Meshal, for all the songs and supportive hugs while I was writing this dissertation, and my daughter, Yara, for cheering me up with all the smiles and cooing sounds. I love you all unconditionally.

I would also like to express my appreciation to all the participants in this project for volunteering to help. I thank King Abdulaziz University for the financial support. I thank George Mason University's Linguistics program and the Digital Scholarship Center for the graduate assistantships over the past five years.

Last but not least, I would like to thank my parents, brothers, and sisters for believing in me and keeping me in their thoughts.

Table of Content

List of Tables

List of Figures

Abstract

THE PRODUCTION AND PERCEPTION OF PROSODIC PROMINENCE IN URBAN NAJDI ARABIC

Hussain Almalki, Ph.D.

George Mason University, 2020

Dissertation Director: Dr. Steven H. Weinberger

This dissertation addresses prosodic prominence in Urban Najdi Arabic (UNA) in both production and perception. Prior research has revealed systematic differences among spoken varieties of Arabic in terms of prominence marking. Further, recent cross-linguistic research shows evidence that acoustic and non-acoustic factors may influence the perception of prominence. Accordingly, this dissertation examines how and to what extent native UNA speakers prosodically mark prominence in different information structures. Additionally, it investigates the influence of acoustic cues and contextual factors on UNA listeners' perception of prominence.

The production experiment examined how UNA speakers utilize acoustic cues to mark information structure, namely focus, and whether they actively disambiguate lexically and propositionally identical utterances according to discourse requirements. The results show that UNA speakers acoustically distinguished different aspects of information structure. The acoustic correlates associated with this acoustic prominence

were duration, maximum intensity, and F0 maximum and F0 range on the stressed syllables of the target words. Speakers used these acoustic cues to disambiguate focus location, focus status, focus size, and focus type. However, speakers did not overwhelmingly distinguish between ambiguous sentences in contrastive and noncontrastive conditions.

The first perception experiment examined how well can listeners perceive acoustic prominence from the speech signal alone. Using data from the production experiment, listeners rated the prominence of sentences by providing prominence ratings for each word on a 5-point rating scale. In this bottom-up design, listeners did not have access to information other than what is in the speech signal. The results show that listeners were highly successful in perceiving prominence based on the speech signal alone. The perceived prominence ratings were in line with the production's findings, in that listeners' ratings of the different aspects of information structure reflected the patterns found in production.

The second perception experiment employed a top-down design to examine whether the perception of prosodic prominence is affected by contextual cues. Listeners read a context question and then listened and rated prominence for each word in the answer on a 5-point rating scale. In this highly controlled experiment, the set-up questions and corresponding answers varied in terms of the question-answer congruence. In the congruent question-answer pairs, the answer was prosodically appropriate to the question. In the incongruent question-answer pairs, the answer was prosodically

inappropriate to the question. Further, the answers in the incongruent cases were identical after different set-up questions to test the independent effect of context.

The findings from the congruent pairs replicated the findings from the bottom-up experiment in that listeners were able to successfully perceive prominence whereby both acoustic and contextual cues complemented each other. The results from the incongruent pairs showed that contextual factors might partially affect the perception of prominence. Specifically, in some conditions of the incongruent pairs, listeners showed evidence of responding to the contextual cues rather than acoustic cues, as reflected by their prominence ratings. However, the effect of the context seems to be phonologically conditioned by the accent distribution on the answer utterance.

Findings from these experiments indicate that the production and perception of prosodic prominence is a multifaceted process that seems to be affected by a combination of acoustic and non-acoustic factors.

Chapter 1

Introduction

The purpose of communication is to share and convey information. Accordingly, speakers generally converse with the intent of conveying information to the listener (or listeners). The speakers' assumption that listeners share some knowledge or information with them guides this communication of information. Therefore, when speakers introduce, for example, *new* information in the context of *shared* or *given* information, they mark that new information as such. Consequently, listeners should be able to retrieve the intended meaning from the way the information is structured. The term *given information* refers to the part of the sentence that has already been made salient in the discourse or is believed to be shared by the interlocutors. Conversely, *new information* refers to the part of the sentence that is new or semantically important to the discourse (Jackendoff, 1972).

Cross-linguistically, there are different ways to mark information structure by manipulating syntax, morphology, and prosody. In Hungarian, for instance, focus is typically signaled by syntactic reordering where the focused item is moved to a pre-verbal position. Accordingly, in Hungarian, the sentence *Peter on the floor slept* would indicate that the word *floor* is focused. If the word *Peter* were to be focused, we would expect a sentence like *on the floor Peter slept* (see Szabolcsi, 1981). In terms of

morphology, some languages indicate focus through a morphological marker; such a language is Guruntum, where the focus is preceded by the morpheme *a* (see Haruna, 2003). It is also possible for one language to use more than one linguistic device to mark focus. This dissertation, however, focuses only on the prosodic aspect of marking the information structure. In this dissertation, the term *prosody* is used as a broad term encompassing intonation, rhythmic patterns, and prosodic phrasing (cf. Selkirk, 2005). Further, the term *prosodic prominence* is defined as a property that makes a word or grouping of words stand out (through acoustic means) relative to other words in the sentence, constituting what is commonly referred to as phrasal stress (or stress beyond the word level) (cf. Terken & Hermes, 2000; Cole et al., 2010).

In non-tonal languages, such as Arabic and English, prosody can serve the crucial function of determining the structure of information for both the speaker and the listener. Prosody can distinguish between important from less important information in a sentence and change the information status of words (e.g., *new* information becomes *given*). For example, the English sentence *John kissed Mary* can convey different kinds of information depending on the context in which it is produced. Consider the following examples in (1) and (2), in which the questions in (1) represent different contexts and the answer in (2) serves as a possible answer to all three questions:

(1) a. Who kissed Mary?

b. Whom did John kiss?

c. What did John do to Mary?

(2) John kissed Mary.

In (1a), the event of "kissing Mary" is made salient (i.e., assumed to be shared by the listener thus is given in the context) and as a result, the answer in (2) would typically emphasize *John* as the *new* information by putting more prosodic prominence on it and *kissed Mary* will be marked as *given* and would typically be deemphasized. In (1b, the event of "John kissing someone" is made salient and, therefore, in the answer, *Mary* would be marked as the *new* information through prosody, and *John kissed* will be marked as the *given* information. In (1c), *John* and *Mary* are made salient; thus, the answer will prosodically mark *kissed* as the *new* information, and *John* and *Mary* will be marked as the *given* information. This structural organization of speech (through prosody in this case) is known as *Information Structure* (e.g., Halliday, 1967; Lambrecht, 1994; Krifka, 2008).

The process of highlighting *new* information in the previous examples is generally referred to as *focus*. *Focus* is an important category of information structure. The information that is highlighted in the discourse, which typically is new to the listener, is said to be focused (Lamdrecht, 1994; Krifka, 2008). As can be inferred from the examples in (1) and (2) above, there is a strong relationship between meaning and information structure, and this relationship seems to be mediated by prosody.

Cross-linguistically, there are open questions pertaining to the relationship between prosody and focus as part of the information structure in both production and perception. For example, although it is generally accepted that focused elements are acoustically more prominent than unfocused elements in the production, questions like what constitutes acoustic prominence in the production and which acoustic features

underlie the listener's perception of prominence are still debated cross-linguistically and within the same language (Eady & Cooper, 1986; Turk & Sawusch, 1997). Additionally, the question of whether contrastively and non-contrastively focused elements belong to distinct categories (Chafe, 1976; Rooth, 1992) and are prosodically differentiated by speakers, and subsequently, perceptually differentiated by listeners is widely debated (Breen et al., 2010; Bishop, 2012). Finally, it is still unclear whether listeners' perception of prominence is signal-driven or context-driven or a combination of both (Cole, Mo, & Hasegawa-Johnson, 2010; Bishop, 2012; Turnbull, Royer, Ito & Speer, 2017). These questions have not been explored for many spoken varieties of Arabic, including the dialect being tested in this dissertation, Urban Najdi Arabic.

To this end, this dissertation examines the interface between prosody and meaning. On the one hand, it examines how prosody, through prosodic prominence, is used to shape the information structure in production to convey different meanings. On the other hand, it examines how this interaction between prosody and information structure influences the perception of the intended meaning. This dissertation uses data from Urban Najdi Arabic. Specifically, this dissertation examines whether and to what extent UNA speakers prosodically mark *focus* as an important information structural category. It also explores the extent to which listeners can retrieve different (focus) information from the speech signal alone in the absence of a context. Finally, it evaluates whether listeners' expectations of how specific information structural categories should be marked influence their perception of *focus* when it appears in a discourse context.

This dissertation advances typological, methodological, and theoretical contributions to the existing literature. First, since the production and perception of prosodic prominence have not been experimentally examined for UNA, this study will add to the existing literature on Arabic by providing new data for a broader typological comparison, both cross-linguistically and among Arabic dialects. Second, methodologically, by including both production and perception experiments, the study provides a comprehensive analysis of different information structures by combining different approaches to speech perception, namely, signal-based analysis (bottom-up) and context-based analysis (top-down). Lastly, this study examines theoretical issues related to both production and perception of prosodic categories, for example, whether contrastive focus and noncontrastive focus are differentiated prosodically and constitute separate categories in UNA. Specifically, findings from signal-based and context-based approaches to perception could have implications for communication theories and speech comprehension. These theoretical investigations will have important implications for the speech-meaning interface in general.

The remainder of this dissertation is structured as follows: Chapter 2 provides a background survey of the relevant literature motivating this study. Chapter 3 constitutes a production experiment investigating how UNA speakers mark focus prosodically and what the acoustic cues associated with prosodic prominence are. Chapter 4 reports on the first perception experiment, which examines whether UNA listeners can perceive intended prosodic prominence from the speech signal alone. Chapter 5 reports on the second perception experiment that investigates whether listeners' expectations of how the

information structure should be marked influence their perception of prosodic prominence. Chapter 6 provides a general discussion of the findings of the three experiments, significant conclusions, and future directions.

Chapter 2

Literature Review

2.1 Information Structure, Common Ground and the Notion of Focus

Information structure (IS) refers how "information units" are structured and organized in an utterance (Halliday, 1967). It is the "packaging" of information to serve a communicative purpose (Chafe, 1976). According to Chafe (1976, p.28), information packaging is a matter of how "the speaker accommodates his speech to temporary states of the addressee's mind" or how the information is delivered. Hence, the speaker's packaging of information is guided by his assumptions or beliefs regarding which pieces of information are shared by the addressee (or listener) and which are not. This shared information or shared knowledge is referred to as Common Ground (CG), a term coined by Stalnaker (1974; 2002). The content of this ground is a combination of both shared or "given" information and "new" information.

On the one hand, given information refers to the information that the speakers believe to be known to the listener (or accepted to be true by the listener). On the other hand, new information refers to the information that the speaker believes to yet be known to the listener (Clark and Haviland, 1977). Accordingly, the CG is continuously updated by the interlocutors as the conversation unfolds, and new information is introduced. It is worth noting that, in addition to adding information to the common ground, sometimes

new information can serve the purpose of altering the common ground in the sense of making a correction (or "repair") to it. That is, one alternative can be switched for another in the common ground (e.g., "No, that's wrong. He did not buy a car. He bought a TRUCK."). However, in this dissertation, I will not investigate repairs.

One of the important notions (or categories) of information structure is *focus*. Krifka (2006) states that "Focus indicates the presence of alternatives that are relevant for the interpretation of linguistic expressions." That is to say, a focused item in an utterance specifies one referent in the presence of a set of alternative referents relevant for the semantic interpretation (Rooth, 1992; Krifka, 2006).

Cross-linguistically, two categories of focus are usually recognized: a) *information focus* (Lambrecht, 1994) and b) *contrastive focus* (Neeleman et al. 2009). According to Lambrecht (1994), information focus includes both broad focus and narrow focus. Broad focus describes the situation where the entire utterance is under the focus domain (Fery, 2007). Unlike broad focus, in narrow focus, only one information unit is highlighted or under the focus domain (Lambrecht, 1994). Contrastive focus describes information units that are in explicit contrast with other information units (Kiss, 1998). The definitions of the different focus types are summarized in Table 1.

Table 1
Definitions and examples of focus types (**bolded** words are focused in the table)

Type	Definition	Examples
Broad focus (BF)	The entire utterance is under the focus domain. This is considered the unmarked or neutral realization of the utterance.	[**Nancy ate the apple**]$_{BF}$ (a.k.a. sentence-focus)
Narrow focus (NF)	A smaller information unit is under the focus domain with no explicit contrast in the discourse.	[Nancy ate] [**the apple**]$_{NF}$ (a.k.a. argument-focus)
Contrastive focus (CF)	A smaller information unit in under the focus domain with explicit contrast introduced in the discourse.	(No,) [**Nancy**]$_{CF}$ [ate the apple] (a.k.a. corrective-focus) answering "Who ate the apple? Mary?"

There are at least three strategies that have been observed, cross-linguistically, for focus marking: *morphological* (through morphological markers specific to focus) (see, Kihm, 1999; Hartmann & Zimmermann, 2009), *syntactic* (through word order or syntactic constructions, like clefting) (see, Birner, 1994; Lambrecht, 2001) and *prosodic* (through prosodic prominence resulting from nuclear pitch accenting and/or phonological rephrasing) (see, Jackendoff, 1972; Ladd, 1996; Buring, 2010). It is worth noting that a language may use more than one strategy in combination to mark focus.

2.2 Focus and the Marking of Prosodic Prominence in Production

It is well established that the status of information structure is fundamental to the prosodic realization of an utterance and that the distribution of prosodic prominence in an utterance reflects differences in how the information units in that utterance are evaluated with regard to the information structure (cf., Halliday, 1967). Cross-linguistically,

prosody is used to express or disambiguate aspects of the information structure, especially in lexically and syntactically identical sentences (Ladd, 2008). For example, in many spoken varieties of Arabic, *yes-no* questions are distinguished from identical declarative utterances through intonation alone (e.g., Egyptian (Hellmuth, 2006; Almalki & Morrill, 2016). The Arabic sentence /li:na na:mat / *Lina has slept* could either be a declarative sentence or a *yes-no* question depending on the way it is produced. A falling intonation signals a declarative sentence interpretation, and a rising intonation signals a *yes-no* question. In many languages, the focus can be marked prosodically in a similar way, as it is the case in English (Ladd, 2008) and Arabic (de Jong & Zawaydeh, 1999).

In practical terms, the prosodic marking of focus can be defined as a communicative function by which a particular information unit is emphasized through prosody (Xu, Chen & Wang, 2012). That is, focused words are highlighted from other words in the utterance by manipulating the acoustic features of these words, including F0, intensity, and duration. Such modulation of the acoustic features determines prosodic prominence. The prosodic prominence of a word is perceptually relative to another word or a phrase (Turnbull et al., 2017). Prosodic prominence is generally equated with accenting (i.e., accented words are more prominent than unaccented ones above the word level). For example, Pierrehumbert & Hirschberg (1990) have found that English speakers tend to assign high pitch accent to words under narrow focus and low-to-high pitch accent to words under contrastive focus.

Cross-linguistically, there are several acoustic features associated with prosodically focused words (hence, accented). These include F0, intensity, and duration.

Essentially, pitch accents are associated with syllables that have higher F0, higher intensity, and longer duration. Focused items have been reported to show longer duration and higher intensity in different languages, such as Greek (Baltazani & Jun, 1999) and English (House & Sityaev, 2003). Additionally, expansion of pitch range, which could be defined as the difference between the minimum and maximum F0, on focused items has been found to be a consistent predictor of prosodic focus across many languages, like English and Mandarin (e.g., Rump & Collier, 1996).

In a recent study, Breen, Fedorenko, Wagner & Gibson (2010) examined whether native English speakers prosodically distinguish different focus types in different locations in SVO sentences. They reported that speakers could prosodically mark focus location (S, V, O), focus breadth (narrow, broad) through intensity, duration, and F0. Additionally, Breen et al. (2010) found that English speakers could successfully distinguish contrastive focus from a noncontrastive focus only when they were made aware of the possible prosodic ambiguity. That is, when speakers were made aware of this distinction between the two possible interpretations, they produced the contrastively focused elements with greater intensity, longer duration, and lower F0 than non-contrastively focused elements.

In addition to the acoustic features discussed above that are specific to a word or information unit under focus, prosodic prominence could further be enhanced through other means. Specifically, the phonetic realization of the surrounding (unfocused) elements can contribute to the prosodic marking of focus. For example, Jun (2014) reports that, in many languages, post-focus items can be de-accented (i.e., they do not

receive a pitch accent and therefore have lower F0 and intensity and shorter durations). Some languages also show compressed pitch ranges (i.e., smaller differences between the maximum and minimum F0) in post-focus items, as in Mandarin (Xu & Xu; 2005). Although pre-focus items do not usually show consistent differences in terms of pitch range, duration, and intensity, there are cases of optional de-accentuation that have been reported for some languages like English (Jun, 2012). The fact that focus effects could extend to influence pre- and post-focal items enhance the prosodic prominence of focused elements in a non-trivial way because it creates a maximal difference between focused and unfocused elements making focused words stand out even more.

2.2.1 Contrastive and Non-Contrastive Focus

The question of whether contrastive and noncontrastive focus are different from each other has long been debated in production studies (e.g., Chafe, 1976; Halliday, 1967; Bolinger, 1961; Rooth, 1992). It is still not well understood whether contrastive focus and noncontrastive focus constitute different categories of the information structure. The existence of a distinction (or lack thereof) between contrastive and noncontrastive focus is a theoretical issue. Many researchers have treated these two as separate categories (e.g., Chafe, 1976; Halliday, 1967, among others), whereas others have treated them as one category and argued that there is no categorical difference between them (e.g., Bolinger, 1961; Rooth, 1992). Therefore, if contrastive focus and noncontrastive focus are indeed two categories (-/+ contrast), then one would expect to find substantial differences in the acoustic realization of the two. Ito, Speer, & Beckman, (2004) have observed such findings for English, where they report that English speakers

usually use a steep rise from a low accent to a high accent to indicate contrastive focus, compared to only a gradual rise to a high accent to indicate noncontrastive focus. Similar results were obtained by Pierrehumbert & Hirschberg (1990) and Breen et al. (2010). However, if there is no difference between the two categories, one would not expect to find any categorical differences (e.g., distinct pitch accents) in the acoustic realization of the two. This, of course, does not rule out the possibility of gradient acoustic differences between the two. The existence of an acoustic distinction between contrastive or noncontrastive focus, or the lack thereof, has implications for both production and perception. However, it is important to note that even if the acoustic distinction between contrastive and noncontrastive focus is not substantial, it does not mean they are not semantically different, but rather, for our purposes, it would simply suggest that there is no perceptible phonetic consequence of such a distinction.

2.3 Focus and the Perception of Prosodic Prominence

Production studies have demonstrated that speakers can prosodically distinguish different information structures, in this case, *focus* and that the acoustic features constituting prosodic prominence include F0, intensity, and duration. However, the factors constituting the perception of prosodic prominence are not well understood. To the best of my knowledge, the perception of prosodic prominence has not been studied for Urban Najdi Arabic. However, because some of the questions asked about perception in this dissertation are theoretical in nature, a cross-linguistic review of relevant perception studies will highlight the important issues at hand.

Since speech perception is linked to the listener's interpretation or comprehension of the intended meaning, at least two factors are thought to influence the perception of prosodic prominence, namely, the acoustic cues in the speech signal (such as F0, intensity, and duration) and the contextual cues (i.e., in the discourse). Although it is well established that listeners can identify prominent words from non-prominent words based on the speech signal alone (without a context) in many languages, it is still unclear whether listeners are always able to distinguish between different types of focus from the signal successfully. In English, for example, Gussenhoven (1983) demonstrates that listeners can perceive differences between utterances produced with broad focus and narrow focus. That is, when listeners rated the verbs in utterances originally produced in broad focus context and utterances originally produced in object narrow focus context, they rated the verbs in the broad focus utterances to be more prominent than the verbs in the utterances where the object was narrowly focused. This finding suggests that speakers prosodically encoded the two focus types differently. It also suggests that listeners were able to successfully perceive and comprehend these differences from the speech signal alone since no context was provided in this experiment. The finding further suggests that the verb *kissed* in the sentence *John kissed Mary* would be realized differently when the sentence is produced as an answer to the question *What happened?* versus *Who(m) did John kiss?* As an answer to the first question, the verb is focused as part of the entire sentence, while as an answer to the second question, it is completely unfocused.

Breen, Fedorenko, Wagner & Gibson (2010) conducted a series of perception studies in English, in which native English listeners matched utterances with appropriate

questions to examine whether listeners could retrieve the intended meaning from the speech signal alone. They used SVO utterances with different focus types in different positions. In the first experiment, they reported that listeners were very successful in identifying focus location (Subject, Verb, Object) but less successful in identifying focus type (noncontrastive vs. contrastive). The fact that listeners were less successful could be because speakers were not always signaling a difference between contrastive and noncontrastive focus in their productions. The researchers reported that speakers acoustically distinguished contrastive from noncontrastive productions only when they were told of the possible ambiguity between the two categories. These results also suggest that listeners may not always be able to retrieve contrast from the acoustic signal since most of the time, there was no perceptible difference in the acoustic signal.

In the second experiment, the researchers added an attribution phrase *I heard that ...* before the target utterances. The results replicated the results from the first experiment with regard to the fact that listeners could accurately identify focus location. Additionally, listeners were more successful in distinguishing contrastive from noncontrastive focus than in the first experiment. They attributed this increase in distinguishing contrastive from noncontrastive to the fact that speakers tended to prosodically mark *I* in the attributive phrase *I heard that* when it is in a contrastive condition. The latter point raises the question of how much additional information outside the target utterance is needed by the listeners to perceive the distinction between contrastive and noncontrastive focus.

Although Gussenhoven (1983) and Breen et al. (2010) have demonstrated that the acoustic cues in the speech signal alone could lead to the successful perception of prosodic prominence, the nature of the interaction between acoustic cues and contextual cues is still not well understood. Specifically, it is not clear yet whether listeners always rely on the speech signal to perceive prosodic prominence regardless of the presence of contextual cues, or if they ignore the acoustic cues and rely solely on the pragmatic/semantic context or if they use a combination of both acoustic cues and contextual cues.

As Turnbull et al. (2017) reasonably argue, there are at least three possible hypotheses to the perception of prosodic prominence. A signal-based hypothesis predicts that the perception of prominence depends on the acoustic saliency of prominent words. That is, contextual cues would only enhance the perception of prominence but cannot override acoustic cues (Bock and Mazzella, 1983). A context-based (or expectation-based) hypothesis predicts that the perception of prominence depends on the context regardless of the actual acoustic saliency of words (Lieberman, 1965). A third hypothesis predicts that both signal-based and context-based cues contribute equally to the perception of prosodic prominence. Under this balanced hypothesis, each factor (signal and context) should contribute equally to the perception of prominence, and in combination, their effect becomes additive (see Cole et al., 2010; Bishop, 2012; Turnbull et al., 2017).

There is a growing interest in examining the interaction between prominence perception and contextual cues, such as a context sentence or context question. A context

sentence (or context question) in a study is typically used to generate a small amount of common ground information against which participants can evaluate a target sentence's information structure with respect to this established common ground. For example, Bock and Mazzella (1983) conducted two comprehension experiments in English, in which target sentences were preceded by context sentences that varied in prosodic appropriateness. They reported that target sentences with contrastively focused subjects (e.g., JOHN kicked the ball) were comprehended faster when preceded by a contextually appropriate sentence in which the subject had the prosodic prominence (e.g., BILL didn't kick the ball) than when it was preceded by a contextually inappropriate sentence (e.g., Bill didn't KICK the ball). Bock et al. (1983) argued that the contextually appropriate context sentence evoked anticipation for a contrastive entity in the target sentence, which in turn made comprehension faster. This finding suggests that the perception of prominence is not completely dependent on the speech-signal but rather that the contextual cues also play a role.

Cole, Mo, and Hasegawa-Johnson (2010) examined the perception of prominence using speech excerpts from the Buckeye Speech Corpus (Pitt et al., 2007). In their experiment, naïve listeners listened to audio stimuli and were asked to mark which words they perceived as prominent on a printed transcript. Since no context was presented to the listeners, the findings show that listeners can perceive prominence from the speech signal alone. However, Cole et al. (2010) also found an effect of expectation-based factors, particularly word frequency and repetition. Specifically, listeners marked relatively unpredictable words (low-frequency words) to be more prominent than predictable words

(high-frequency words), perhaps because they required more processing effort. Although this study did not control for the context, the two findings show that prominence perception can be both signal-driven or expectation-driven.

Bishop (2012) conducted two perception experiments to explore English listeners' knowledge of how the size of a focused constituent is expressed prosodically (e.g., object focus in which only the object is focused vs. predicate focus in which both the verb and object are focused). The goal was to find out how listeners' expectations of how different information structures are (or should be) produced might influence their ratings of prosodic prominence. The idea was that since speakers are expected to produce a focused object in a broad SVO sentence with greater prominence, listeners would have the same expectations about how the information structure should be marked when they listen.

In the first perception experiment, Bishop recorded two native speakers of American English, producing the data in sets of question-answer exchanges. Each question-answer pair had three experimental conditions: one with the entire sentence under focus (broad focus), one with narrow focus on the verb, and one with narrow focus on the object. The question-answer pairs in the examples in (3) below illustrate the experimental conditions:

(3) a. What happened yesterday?

b. What did you do yesterday?

c. What did you buy yesterday?

d. I **bought a motorcycle**.

Readers were instructed to be as natural as possible in their readings. Because the purpose of the study was to test the independent effect of information structure on the perception of prominence for words or information units in the answer sentences, the recordings of answer sentences produced in response to VP focus questions in the original recordings were extracted and used as the answer to the three focus conditions. That is, listeners would hear three different questions eliciting different focus conditions for each question-answer exchange but will hear the same answer after each question.

Listeners were asked to listen to question-answer exchange and rate prominence of the verbs and objects in the answer utterances on a scale of 1-5, where 5 = very prominent. The results showed that listeners used context (in this case, the set-up question) to judge prosodic prominence. Specifically, listeners heard words in the object position to be more prominent when the question was asking about the object, which makes it narrowly focused, than when it was in the context of broader focus. Furthermore, the perception of more prominence on the object was made relative to the verb. That is, objects were not simply heard as more prominent, but also that verbs were heard as less prominent. This finding suggests that listeners' expectations of how the information structure should be marked affected their judgments because what they rated to be prominent was not prominent in the actual signal. That is, this pattern is only expected if listeners have clear expectations about what speakers do, and this expectation or knowledge affected their perception.

These results demonstrated that listeners could not have been responding to something in the speech signal because they heard the same auditory stimulus. It is still

unclear from these results if listeners were completely ignoring the acoustic cues in the speech signal. Bishop (2012) speculates that perhaps listeners were initially attending to the signal, but their expectations about how information structures should be realized in these contexts ended up modulating their judgments. This is because there was no difference in the speech signal between the utterances.

In the second experiment, Bishop (2012) tested the effect of focus size for contrastive focus in three focus conditions: entire sentence, verb phrase, or object. Question-answer pairs, which included complementizer phrases headed by *because* were used to elicit contrast. The examples in (4) illustrate these conditions:

(4) a. Why's your wife mad? Because your roof's leaking?

b. Why's your wife mad? Because you lost your job?

c. Why's your wife mad? Because you bought a car?

d. No… because I **bought a motorcycle**.

Following the first experiment, the listeners heard different questions and the answers were the sentences where the verb phrases were contrastively focused. Therefore, listeners would hear questions representing different focus conditions but hear the same answer. Like in the first experiment, native speakers of English were asked to listen to question-answer dialogues corresponding to different information structures and assign prominence ratings to the verbs and objects in the answers.

These findings in Bishop (2012) showed that contrastively focused objects were perceived to be more prominent than objects in broader focus constituents. Like the first experiment, in addition to rating contrastively focused objects as more prominent verbs

were perceived to be less prominent when the object was focused. Additionally, when the question was targeting contrastive focus on the verb phrase, listeners successfully rated both the verb and object to be more prominent (than their ratings of the verb alone or object alone under the other focus conditions). This finding indicates that listeners' ratings were highly correlated with aspects of the signal since the utterance they heard had the verb phrase contrastively focused (i.e., matching the context). It suggests that when the listener's expectations of how information structure should be marked match the acoustic information in the speech signal, it enhances the perception of prosodic prominence. It is clear from Bishop's (2012) study that the context questions alone can lead to differences in the perception of "illusory" prominence in the target utterances.

However, there are still open questions related to whether listeners balance the effects of acoustic and contextual cues when rating prominence. Specifically, it is still unknown whether these contextual cues or acoustic cues are additive, or if one could "override" the other, especially when listeners are presented with conflicting information between the context and auditory stimulus. Although it seems that listeners generally have certain expectations about how speakers should produce different types and sizes of focus, it is not clear whether listeners' expectations alone could completely override their perception of the cues in the speech signal.

This dissertation deals with issues pertaining to prosodic prominence in both production and perception in Urban Najdi Arabic (UNA). Since the issues discussed thus far have not been widely tested for Arabic in general and have not been explored for UNA, in particular, a typological overview is provided in the following section.

2.4 Typological Overview of Arabic Prosody

2.4.1 Focus in Arabic: Syntactic and Prosodic Expression of Focus in Arabic

Arabic is characterized as having relatively free word order (Bakir, 1979), and thus, has the option to express focus either syntactically or prosodically. Moutaouakil (1989) examines the information structure of Modern Standard Arabic, including focus. He proposes VSO as the canonical word order in MSA. He recognizes several IS categories in MSA, two of which are relevant here: narrow focus (information focus in Moutaouakil's term) and contrastive focus. According to Moutaouakil, narrow focus can only be expressed in-situ (example 5a). He uses the idea of "new information" to describe narrow focus. Accordingly, narrow focus cannot be syntactically distinguished from broad focus in MSA. Contrastive focus, however, can be expressed syntactically in three ways: It can be expressed, ex-situ in the left periphery (5b), pseudo-clefting (5c), or by negative-restrictive construction (5d):

(5) a. /akaltu [tamran]/
ate-1sg tamran-acc
'I ate [$_{\text{DATES}}$]'= in response to "what did you eat?".
b./[Ali-an] qabaltu/
Ali-acc met-1sg
'It was [$_{\text{ALI}}$] that I met'= in response to "who did you meet? Zayd?".
c. /allathi qabal-tu-hu [Ali-an]/
the-one that met-1sg-him-3sg Ali-acc
'It was [$_{\text{ALI}}$] that I met'= in response to "who did you meet? Zayd?".
d./ma qabal-tu illa [Ali-an]/
not met -1sg but Ali-acc
'I met only $_{\text{[ALI]}}$' in response to "who did you meet? Zayd?".

Focus and information structure have also been studied in some spoken varieties of Arabic. In spoken Bahraini Arabic, Holes (2010) makes the distinction between new and given information and claims that new information focus can be expressed either prosodically or syntactically (ex-situ in the left periphery or pseudo-clefting. Old information can either be expressed ex-situ in the left or right periphery. Holes (2010) does not discuss the idea of narrow versus contrastive focus, so it is not clear how they are expressed in this dialect.

Ingham (2010) describes some aspects of the information structure of Bedouin Najdi Arabic, a variety spoken in northern Saudi Arabia. Similar to Holes (2010), Ingham uses the notion of new and given information. He reports that new information can be expressed prosodically or syntactically ex-situ in the left periphery. Old information is expressed ex-situ in the left or right periphery.

Alzaidi (2014) examines how information focus (narrow and broad) and contrastive focus are encoded prosodically in Hijazi Arabic (HA), a dialect spoken in the western part of Saudi Arabia. He also examines whether word order or intonation is useful in identifying focus. He shows that information focus is obligatorily realized in-situ in the syntax, but contrastive focus can be expressed in-situ or by left-dislocation. Evidence showing how MSA expresses contrastive focus (Moutaouakil, 1989) and the fact that contrastive focus in HA is optionally expressed ex-situ supports the distinction between contrastive focus and narrow focus as separate categories. This distinction seems to be syntactic/semantic. As discussed in the following section, Alzaidi (2014) shows that intonation and not word order is crucial for focus marking in HA utterances. To

summarize, previous studies on Arabic indicate that focus can be expressed syntactically or prosodically in different varieties of Arabic. The syntactic marking of focus seems to be optional and restricted to a certain focus type (i.e., contrastive focus).

Although studies dealing with the realization of prosodic focus in different spoken varieties of Arabic is discussed in the next section, previous studies seem to suggest that prosody, not syntax, is crucial for focus marking as in Hijazi, Lebanese, and Egyptian Arabic (Alzaidi, 2014; Chahal and Hellmuth, 2014). Accordingly, the question of whether there are phonetic differences between the realization of narrow focus and contrastive focus within each of these Arabic varieties is still open.

This dissertation asks whether and how prosody contributes to the marking of focus in SVO sentences in UNA under different focus conditions. In answering these questions, the phonetic realizations of both narrow and contrastive focus are directly compared. Evidence from previous literature on modern Arabic dialects points to the fact that SVO is considered the canonical and unmarked word order in these modern dialects (Soltan, 2007; Musabhien, 2009, Alshamari, 2017), and thus, is assumed for UNA and consequently used in the stimuli as will be described in the method section.

2.4.2 Production of Prosodic Prominence in Arabic

There is increasing evidence from some spoken varieties of Arabic (e.g., Lebanese, Egyptian, and Hijazi) showing that focus is marked prosodically and that this prosodic marking influences not only the focused item but also the other items in the utterance. In these spoken varieties of Arabic, F0, intensity, and duration have been found

to distinguish focused from non-focused items. However, these spoken varieties seem to have some variability among them (Chahal, 2001; Hellmuth, 2006; Alzaidi, 2014).

Chahal (2001), for example, examined the prosodic marking of broad and narrow focus in Lebanese Arabic. She found that the phonetic realization of the focus is determined by the position in which the focused items occur. That is, focused words in the initial position displayed different acoustic specifications in terms of F0, intensity, and duration than focused words in the medial or final position. In terms of the overall intonational pattern, for example, Chahal (2001) reported that focused items always received the nuclear pitch accent regardless of the position of the focused items. Pre-focused items are optionally de-accented, and post-focus items were always de-accented. Chahal (2001) also acoustically compared narrowly focused utterances with broad focus counterparts and found that narrow focused items were produced with expanded pitch range, higher F0, and higher intensity, whereas broad focus utterances showed a default neutral intonational pattern. In broad focus utterances, the final words were always the most prominent and received the nuclear pitch accent. Finally, Chahal (2001) notes that pre- and post-focused items reflect pitch range compression, which in turn contributed to the marking of prosodic prominence.

Hellmuth (2006) examined focus in Egyptian Arabic and found that in terms of the overall intonational patterns, there is no evidence of de-accentuation on unfocused items (i.e., given information) and every content word received a pitch accent. Further, Hellmuth (2006) reported that contrastive focus is marked by expanding the F0 range on

the focused item and compression of the F0 on post-focus items. De-phrasing and utterance final focus lengthening were optionally used to mark focus in Egyptian Arabic.

In Hijazi Arabic, Alzaidi (2014) reported that the nuclear pitch accent falls on the focused item, and that pitch is compressed on post-focus items. Further, focused words showed an expanded pitch range. Post-focus words also showed lower F0 values. Alzaidi (2014) concluded that maximum F0 and pitch range were the most important focus correlates in Hijazi Arabic.

Evidence emerging from these studies suggests that spoken Arabic varieties vary in the way they mark focus prosodically. Moreover, it shows that focus marking affects not only the focused items but possibly the pre- and post-focused items. Generally, the prosodic correlates to prominence in spoken varieties of Arabic include the manipulation of F0, intensity, and duration on the focused and the adjacent non-focused items. In terms of overall intonational patterns, these varieties are also different from one another.

Since the relationship between prosody and information structure is experimentally understudied for UNA, Almalki (2016) designed a production experiment in the form of question-answer pairs examining different focus types on noun phrases (broad, narrow and contrastive) in different utterance positions (initial, medial and final). The experimental utterances had "subject-verb-object-preposition-object of preposition" word order adopted from Chahal (2001) (see example in Table 2). The purpose of the study was to determine whether and how focus is marked prosodically in UNA and to explore whether there are prosodic differences between these focus types in different sentential positions. Eight native speakers of UNA produced 336 SVO utterances in three

focus conditions (broad, narrow, contrastive) and three positions (subject, direct object, object of preposition).

The data were acoustically analyzed in terms of duration, F0, and intensity. The results showed that these acoustic cues are crucial to the marking of focus in UNA. The analyses revealed that broad focus is different from both narrow and contrastive focus in all the examined acoustic aspects. Focused items showed longer duration, higher F0 maximum, lower minimum, wider range, and higher intensity. Post-focus items were dominantly de-accented. The de-accentuation was realized by reduced duration, lower F0 values, compressed F0 range, and lower intensity. Finally, no acoustic difference was found between narrow and contrastive focus.

Table 2
Almalki's (2016) sample experimental sentences. (test words are under subject, direct object (D.O) and object of preposition (Obj of prep.) columns. English translation and IPA transcription are provided).

		Subject	Verb	D.O	Preposition	Obj of Prep.
1	IPA →	ri.ma	ʃa.fat	sa.rah	maʕ	nu.ra
	Translation →	Rima	saw	Sarah	with	Nora
2	IPA →	lu.si	ħa.mat	li.ma	min	ra.mi
	Translation →	Lusy	protected	Lima	from	Rami

Taken together, these studies demonstrate that although these spoken Arabic varieties seem to be similar in the sense that they use F0, intensity, and duration to mark focus, the specific nature of phonetically realizing these correlates is different among them. Additionally, previous findings encourage analyzing as many acoustic cues as possible rather than focusing on, for example, F0 alone. Further, the lack of a clear

distinction between contrastive and noncontrastive focus observed for Urban Najdi Arabic (henceforth UNA) calls for further investigation. Additionally, Almalki (2016) did not include an acoustic analysis of focus condition on the verb phrase nor the verb, which will be considered in the current dissertation. Finally, in Almalki's (2016) data, the items in the object of preposition position were in the utterance-final position, and this might have resulted in inaccurate description of the data since focus in utterance final position might have been influenced by acoustic factors specific to the final position, such as lower F0, lower intensity and utterance-final lengthening.

2.5 Dissertation Research Questions

Guided by previous findings on the production and perception of prosodic prominence, this dissertation asks the following questions about Urban Najdi Arabic (UNA):

1. Do speakers of UNA mark information structure prosodically in SVO utterances, and to what extent?

It is predicted that UNA speakers will prosodically distinguish different information structures. Specifically, it is predicted that the size of the focus (entire sentence, a phrase, or a word), the location of the focus (subject, verb, object) will be distinguished prosodically. In terms of focus type, it is predicted that broad focus will be different from narrow focus. However, given previous findings from UNA, it is expected that narrow and contrastive focus will not be prosodically differentiated from each other.

2. What are the acoustic features associated with the prosodic prominence of the different information structures in SVO utterances in UNA?

It is predicted that prosodic prominence will be associated with increased duration and intensity and higher F0 values on the focused word and decreased duration and intensity and lower F0 values on unfocused words. Manipulations of these acoustic features are considered typical predictors of prosodic prominence in many spoken Arabic varieties, including UNA, as well as in many pitch accent languages, like English and German.

3. In the absence of a discourse context, how well can naïve UNA listeners retrieve different aspects of the information structure intended by the speaker from the speech signal?

It is expected that naïve UNA listeners will be able to distinguish what speakers encode in their speech successfully. That is, listeners are predicted to be able to perceive prosodic prominence, signaling different sizes of focus (entire sentence, a phrase, a word), focus locations (subject, verb, object), and focus types (broad and narrow).

4. Does naïve UNA listeners' knowledge of how different aspects of the information structure are marked prosodically influence their perception of prosodic prominence?

It is hypothesized that if listeners' knowledge or expectations of how different aspects of the information structure are marked prosodically influences their perception of prosodic prominence, their perception will be guided by the context even when the speech signal demands otherwise. In this case, listeners are expected to ignore the speech signal and hear "illusory" prominence created by their expectations. It is, however, also possible that context does not override what is in the speech signal, and listeners will be responding to what is in the signal regardless of the context. In this case, their performance is expected

to be similar to the performance from the first perception experiment. A third possibility is that both what is in the speech signal and the context are important to the perception of prosodic prominence. In this case, one would expect that listeners will hear some "illusory" prominence matching the context when the context conflicts with the speech signal, but at the same time, they will hear an even stronger prominence when the context and signal are matching.

Chapter 3

Production Experiment

This chapter reports on a production experiment that elicited short SVO utterances in different focus conditions from UNA speakers. Although Arabic has a relatively flexible syntax, SVO is considered the canonical word order in spoken varieties of Arabic, including UNA (Soltan, 2007; Musabhien, 2009; Alshamari, 2017). In this experiment, participants listened to pre-recorded questions and read corresponding simple SVO sentences as the answers in a question-answer paradigm. Having participants read the answers made it possible to control for the phonetic environment and word position in the utterance (i.e., subject, verb, and object), among other things, as will be described. Since duration, F0, and intensity have been reported to be important phonetic correlates of prosodic prominence cross-linguistically, as well as for many varieties of Arabic, these acoustic correlates were measured and analyzed in the data analysis.

3.1 Methods

3.1.1 Participants

Eight native speakers of UNA (4 females and four males, age: 26–33 years) participated in this production experiment. All participants had grown up in Riyadh, Saudi Arabia, and were students at George Mason University at the time of the recording. All participants were speakers of English as a second language and reported no speech or

hearing problems. All participants were able to read Arabic and had normal or corrected to normal vision. The demographic information of the speakers is summarized in Table 3 below.

Table 3
Demographic information of the speakers

Gender	*N*	Age in years			Age of the L2 onset in years			Length of residence in the US in years		
		m	s.d.	range	*m*	s.d.	range	*m*	s.d.	range
Female	4	31	2.12	28-33	10	4.69	2-14	2.25	1.09	1-4
Male	4	27	1.73	26-30	10.5	2.60	6-14	2.5	0.86	2-4

3.1.2 Material

The experiment was a reading task in the form of question-answer pairs, in which participants listened to a set of pre-recorded questions in UNA produced by a native speaker and read the corresponding answers on a computer screen. This design was used to control for word order, focus condition, utterance length, as well as syllable shape.

To elicit utterances with different focus conditions, a Subject-Verb-Object-Adverb paradigm was adopted from Chahal (2001). Under this paradigm, target words were in one of the following positions: the subject, verb, predicate, or object position. The final adverbial phrase was included in all the sentences to avoid possible boundary-adjacent effects (such as utterance-final lengthening) on the last word of the sentence (Hirotani, Frazier, and Rayner; 2006). Six experimental sentences yielded a total of 18

target words: 6 in the subject position, 6 in the verb position, and 6 in the object position (See Table 4 and Table 5). For predicate condition, the target words are the words in the verb and object position. All target words are disyllabic, with the lexical stress falling on the first syllable.

Table 4
Production experimental sentences. (the target words are under subject, verb, and direct object columns. IPA transcription and English translation are provided, and stress-bearing syllables are underlined).

		Subject	**Verb**	**Object**	**Adverb**
1	IPA →	θa:mir	ha:waʃ	sa:rah	ʔilju:m
	Translation →	Thamer	scolded	Sarah	today
2	IPA →	sa:lim	sa:maħ	ri:ma	ʔilba:riħ
	Translation →	Salem	forgave	Rima	last night
3	IPA →	ʕa:mir	xa:sˁam	di:na:	ʔilju:m
	Translation →	Amer	argued (with)	Dina	today
4	IPA →	sa:mi:	ra:sal	li:na	ʔams
	Translation →	Sami	texted	Lina	yesterday
5	IPA →	na:jif	ma:zaħ	ja:ra	ʔilba:riħ
	Translation →	Naif	joked (with)	Yara	last night
6	IPA →	ra:mi:	dʒa:mal	nu:rah	ʔilju:m
	Translation →	Rami	complemented	Norah	today

As shown in Table 4, each target word has two syllables, and the first syllable carries the lexical stress. The vowel on the first syllable of each target word is consistently a long vowel. Since we are concerned with the stressed syllable (hence V1), vowel length was controlled so that all words had long vowels. The target words in the subject and object position are all proper nouns (male names for subjects, female names for direct objects). Additionally, the verbs are always in the simple past tense. The

stressed vowels are always followed by a non-stop consonant (which is generally a sonorant, except for three fricatives).

For each of the six sentences in Table 4, 9 elicitation questions were asked to elicit the following productions: 1) a broad focus sentence, where the focus is on the entire sentence, 2) a narrow focus sentence, where the subject is focused, 3) a narrow focus sentence, where the verb is focused, 4) a narrow focus sentence, where the object is focused, 5) a narrow focus sentence, where the predicate is focused, 6) a contrastive focus sentence, where the subject is contrastively focused, 7) a contrastive focus sentence, where the verb is contrastively focused, 8) a contrastive focus sentence, where the object is contrastively focused and 9) a contrastive focus sentence, where the predicate is contrastively focused. This resulted in 54 question-answer pairs. Table 5 below illustrates the different conditions for one of the target sentences (see Appendix A for a complete list):

Table 5
Target focus conditions illustrated by question-answer pairs for one sentence

	Condition	Elicitation Question	Target answer
1	broad focus	**wiʃ sˤar ʔilju:m** what happened?	
2	narrow focus [s]	**mi:n** ha:waʃ sa:rah ʔilju:m who scolded Sarah today?	
3	narrow focus [v]	θa:mir **wiʃ sawa** ʔil sa:rah ʔilju:m what did Thamer do to Sarah today?	(laʔ) θa:mir ha:waʃ sa:rah ʔilju:m
4	narrow focus [o]	θa:mir ha:waʃ **mi:n** ʔilju:m Whom did Thamer scold today?	(no) Thamer scolded Sarah today
5	narrow focus [vp]	θa:mir **wiʃ sawa** ʔilju:m What did Thamer do today?	
6	contrastive focus [s]	**mi:n** ha:waʃ sa:rah ʔilju:m xali:d who scolded Sarah today? Khaled?	
7	contrastive focus [o]	θa:mir **ha:waʃ mi:n** ʔilju:m danah Whom did Thamer scold today? Danah?	
8	contrastive focus [v]	θa:mir **wiʃ sawa** ʔil sa:rah ʔilju:m samaħha What did Thamer do to Sarah today? forgave her?	
9	contrastive focus [vp]	θa:mir **wiʃ sawa** ʔilju:m wasˤal ɣadah What did Thamer do today? dropped off Ghadah?	

The elicitation question is what participants heard. The target answer is the corresponding answer to that question. Each question-answer pair represents one focus type in one utterance position. In addition to the experimental sentences, 60 question-answer pairs were used as fillers. The fillers appeared in similar conditions as the

experimental pairs. Still, they differed from the experimental sentences in terms of their syntactic structure (e.g., /ali: illi sa:far ʔiljum/ “Ali is the one who traveled today”).

3.1.3 Procedure

Participants were seated in a sound-attenuated booth in front of a laptop screen (Apple 11.6 MacBook Air) in the Acoustics Lab at George Mason University. Before they began the experiment, participants were given the consent form to read and completed a brief demographic information questionnaire (see Appendix B). Once they completed these steps, the experiment was presented to them using PsychoPy software (Peirce, 2007).

In the experiment, participants listened to pre-recorded questions by a native UNA speaker, one question at a time. After they listened to the question, the corresponding answer to that question appeared in the center of the screen in black Arabic orthography with a white background. Participants were instructed to read the corresponding answer out loud, as naturally as possible. Their reading was audio recorded using the MiC-Apogee microphone plugged directly into the computer using PRAAT program (Boersma & Weenink, 2015).

At the beginning of the experiment, participants completed a short training session to familiarize themselves with the task. The training session comprised of 6 pre-recorded questions and their written answers. The question-answer pairs resembled the structure of the experimental session. However, none of the sentences used in the training session were included in the actual experiment. The total number of experimental question-answer pairs was 54 (6 sentences x 9 questions for each sentence), resulting in

54 unique productions from each speaker. The 54 experimental question-answer pairs and the 60 question-answer fillers were fully randomized and presented in one block. The experiment took, on average, 25 minutes to complete.

3.1.4 Data Analysis

The audio recordings were analyzed using PRAAT (Boersma & Weenink, 2015). For the acoustic analysis, each target utterance was extracted and saved as a WAV file. This resulted in 432 target sound files included in the analysis (54 utterances x 8 speakers). The researcher, who is a trained phonetician/labeler and a native speaker of UNA, selected and labeled each vowel in each word as V1 or V2 depending on whether the vowel was in the first or second syllable. The vowels were selected from the offset of the preceding consonant to the onset of the following consonant as determined by listening and examining the F1 and F2 formant bars on the spectrogram. Vowels rather than the entire syllables in which they occurred were favored because previous studies on UNA (Almalki & Morrill, 2016; Almalki, 2016) consistently revealed that the locus of the phonetic change occurred on the stressed vowel.

Exploratory data analysis revealed that the duration of the unstressed vowel (V2) in the focused words was not affected by focus, as the differences in terms of duration between unstressed vowels in the focused and unfocused conditions were not statistically significant in an analysis of the variance test ($F(1) = 3.08$, $p = 0.08$). The locality of the phonetic change to the stressed vowel was also reported for other spoken varieties of Arabic (Alzaidi, 2014) and American English (Cole et al., 2010). Following common practices in the field, after the data were labeled by the author, 86 utterances, which

represent 20% of the data, were randomly selected and checked for labeling accuracy by a second trained phonetician (see Gut & Bayerl, 2004).

Using a PRAAT script, measurements were taken for the duration (in milliseconds), maximum intensity (in dB), and minimum and maximum F0 in Hz from each stressed vowel. The F0 values were transformed to semitones relative to 100 Hz. The amount of F0 change (i.e., F0 range) within a vowel was also calculated by subtracting the minimum F0 from the maximum F0 within that vowel using the transformed semitones values.

3.2 Results

Four hundred and thirty-two utterances were analyzed, including a total of 1296 vowels (432 utterances x 3 stressed vowels in each utterance). Values of duration, F0 maximum, F0 minimum, and maximum intensity, were taken from the stressed vowel of every target word. The values for duration, maximum intensity, F0 maximum in semitones, and the calculated F0 range in semitones were used in the statistical analysis.

The purpose of the statistical analysis was to determine if these acoustic cues predict information structure categories, focus location, and focus status. In other words, whether there is a statistically significant relationship between these acoustic cues and aspects of the information structure. Accordingly, these measurements were modeled as dependent variables in a series of mixed-effects regression models implemented in the program R (Baayen et al., 2008; R Core Team, 2014), using the lmer() function of the lme4 package (Bates, Maechler & Bolker, 2013).

Models with different independent (or predictor) variables and random effects were constructed and compared with likelihood ratio tests using the ANOVA function. Variables that significantly improved model fit were retained in the best-fitting models for each measurement. For these models, the independent variables always included focus type (broad, narrow, contrastive), focus status (focused, unfocused), and item location (subject, verb, object). Models with F0 measurements also included the gender of the speaker (male, female) as a predictor. Models with maximal random effects structure failed to converge for all of the measurements reported below; therefore, a forward best-path method was used to determine which random slopes to include (Barr, Levy, Scheepers & Tily, 2013). Only focus status met the inclusion criterion (alph = .2), and thus was included as a random slope with random intercepts for participants and items. This is true for all of the measurements, except for the F0 range, which included only the random intercepts for participants and items. The results from the best model are shown for each measure (in their respective sections below) with parameter specific *p*-values obtained by using the Satterthwaite approximation, implemented in the lmerTest package in R (Kuznetsova et al., 2017).

3.2.1 Duration

This section reports on the results from the duration (measured in milliseconds.) of the stressed vowel in each target word, which in our data was always the first vowel of a disyllabic word. Table 6 summarizes the descriptive statistics for duration arranged by focus condition, focus type, and item location with the standard deviation for each condition. These results are also visualized in Figure 1 below.

Table 6
Target words' average stressed vowel duration (ms)

Focus Type	Item Location	Focus Status			
		Focused		Unfocused	
		M	SD	M	SD
Broad	*Subject*	109	27.7	NA	NA
	Verb	98.4	19.6	NA	NA
	Object	108	23.2	NA	NA
Narrow	*Subject*	140	21.2	99.9	22.1
	Verb	114	24.6	95.4	18.3
	Object	117	23.1	100	22.3
Contrastive	*Subject*	139	28.9	103	24.1
	Verb	116	25.5	94.3	20
	Object	118	22.8	97.4	22.1

Overall, words that are focused have longer vowel durations (M = 117, SD = 26.3) than unfocused counterparts (M = 98.8, SD = 21.3). Additionally, words that are under narrow (M = 120, SD = 25.3) and contrastive focus (M = 121, SD = 26.5) have longer durations than words under broad focus (M = 105, SD = 24). This pattern is expected to be found as focus have been found to affect duration in previous cross-linguistic research. These observations can be further explored in Figure 1 below.

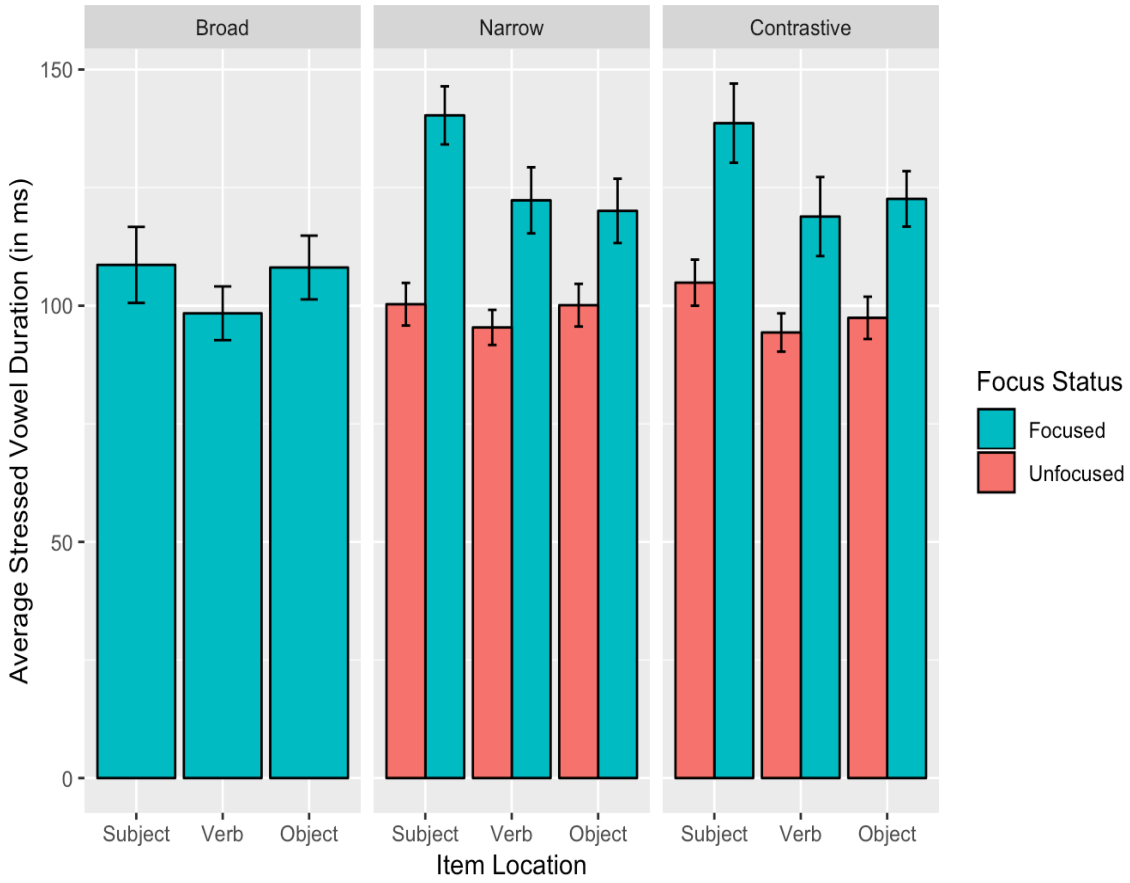

Figure 1 Target words' average stressed vowel duration (ms.) (error bars represent 95% confidence interval).

As can be seen from Figure 1, narrowly and contrastively focused words show longer stressed vowel duration than broadly focused words, which in turn show longer duration than unfocused words in narrow and contrastive conditions. Further, the figure shows a strikingly similar pattern between narrow and contrastive focus conditions.

To test these patterns, a series of mixed-effects regression models were constructed in R (R Core Team, 2013) using the lmer() function of the lme4 package (Bates, Maechler & Bolker, 2013). The independent variables were focus type (broad, narrow, contrastive),

focus status (focused, unfocused), and item location (subject, verb, object). The focus type variable represents the focus condition of the sentence, either broad, or narrow, or contrastive. The focus status variable represents the status of each word in the sentence as being focused or not focused. The item location variable represents whether the word is in the subject, verb, or object position independent from its focus condition. The interactions of these variables give us the nine possible conditions described in the methods section. Focus status was included as a random slope with random intercepts for participants and items. The results from the best mixed-effects model are shown in Table 7 with parameter estimate β statistics, standard error, t value, and p-value for the fixed effects.

Table 7
Coefficients of the best linear mixed-effects model of duration (ms) (N= 1296).

	β	SE	*t*	*Pr(>|t|)*
Fixed effects				
(Intercept)	108.65	4.11	26.43	**<0.001**
Focus Type (Contrastive)	-0.20	1.24	-0.16	0.87
Focus Type (Broad)	-18.99	1.94	-9.81	**<0.001**
Focus Status (Focused)	25.60	3.35	7.65	**<0.001**
Item Location (Verb)	-4.09	1.39	-2.96	**<0.01**
Item Location (Subject)	11.69	1.24	9.42	**<0.001**
Contrastive * Focused	-0.58	2.50	-0.23	0.81
Contrastive * Verb	-0.89	2.90	-0.31	0.76
Broad * Verb	-7.01	4.59	-1.53	0.12
Contrastive * Subject	-1.23	2.74	-0.45	0.65
Broad * Subject	-17.76	4.23	-4.20	**<0.001**
Focused * Verb	1.21	2.90	0.42	0.67
Focused * Subject	18.31	2.78	6.59	**<0.001**
Contrastive * Focused * Verb	1.49	5.81	0.26	0.79
Contrastive * Focused * Subject	8.60	5.56	1.55	0.12
	s^2			
Random effects				
Participant (Intercept)	85.83			
Focus Status (Slope)	77.02			
Item (Intercept)	34.30			

Model comparisons were conducted to test the significance of these factors and their interactions. This was done by comparing the full model (the model with the best fit described above) against the model without the effect in question. Accordingly, models with different predictors were computed and compared with likelihood ratio tests carried out by the ANOVA function, which used log-likelihood as a measure of goodness of fit (cf. Baayen, 2008).

First, three models were constructed and compared against the full model to test the main effects. The first model was identical to the full model, except it did not include

focus type as a main effect. The comparison revealed that focus type significantly affected the duration of the stressed vowel ($\chi^2(1) = 107.67$, $p < 0.001$). By looking at the coefficient table above, the source of this significant effect seems to be coming from the fact that narrow focus and contrastive focus had significantly longer durations of the stressed vowel than broad focus. The second model was identical to the full model, except it did not include the focus status as a main effect. The comparison revealed that focus status significantly affected the duration of the stressed vowel ($\chi^2(1) = 60.81$, $p < 0.001$), reflecting that focused items had significantly longer vowel duration than unfocused counterparts. The third model was identical to the full model, except it did not include item location as a main effect. The model comparison revealed a significant main effect of item location ($\chi^2(1) = 120.84$, $p < 0.001$), reflecting that duration significantly decreased as item location moved from the subject to the object.

Additionally, two more models were constructed to test the interaction between these factors. The first model testing the interaction was identical to the full model, except it did not include the interaction between the focus type and item location. The comparison revealed a significant interaction between the two factors (i.e., focus type and item location) ($\chi^2(1) = 21.06$, $p < 0.001$), indicating that subjects, verbs, and objects have longer durations in narrow and contrastive focus compared to words in the same position in broad focus utterances, as can be seen from the coefficient table above. The second model testing interactions was identical to the full model, except it did not include the interaction terms between focus status and item location. The comparison revealed that the interaction between focus status and item location is a significant one ($\chi^2(1) = 43.18$,

$p < 0.001$), indicating that the duration pattern for item location is different between focused and unfocused words.

Further follow-up comparisons were conducted using contrast coding to confirm the findings obtained for focus types, specifically to explore whether there are differences between narrow and contrastive focus. First, in order to test whether there is a statistically significant difference between narrow and contrastive focus, broad focus was excluded from the dataset, then compared the same full model described above against a model without focus type. That comparison revealed that focus type (without broad focus in this case) did not contribute significantly to the model fit ($\chi^2(1) = 1.84$, $p = 0.87$), indicating that the difference in terms of the duration of stressed vowels between narrow focus and contrastive focus was not significant. To compare the differences between broad focus and focused words under narrow and contrastive focus, words that were coded as focused in the focus status variable were included. Then, the full model was compared against the model without the focus type variable. The comparison revealed that focus type contributed significantly to the model fit ($\chi^2(1) = 132.67$, $p < 0.001$), indicating that the duration of the stressed vowels in focused words under narrow and contrastive focus were significantly longer than those under broad focus.

Additionally, another model was constructed to test whether the duration of stressed vowels in words under broad focus is different from that of unfocused words. Accordingly, a subset of the data that only included the duration for broad focus words (whose focus status is focused) and the durations for the unfocused words from narrow and contrastive focus (whose focus status is unfocused) was created. Two models were

then compared, the one with focus status against the one without focus status. The comparison revealed a significant effect of focus status ($\chi^2(1) = 22.14$, $p < 0.01$), indicating a significant difference between broad focus and unfocused words under narrow and contrastive focus. Accordingly, it was found that duration was different between broad focus and the other focus types, as well as between focused and unfocused items. Broad focus shows durations that are significantly shorter than items under narrow or contrastive focus and significantly longer duration than the unfocused items. This pattern is expected, since under broad focus, every word is relatively of equal importance. In contrast, under other focus types, the phonetic cues are expected to be maximized on the focused items and minimized on the unfocused items.

Additionally, in order to test whether focus location (i.e., focus is on the verb, the object or the entire predicate) has a consequence on the duration of the stressed vowels in these items, a subset of the data that only included narrow and contrastive focus for the focus type and verb, object, and predicate for the focus location variable and verb and object for the item location was created. A new full model for this dataset was created where item location, focus location, and focus type, as well as interaction between item location and focus location, were entered as independent variables with participant and sentence are random effects. This new full model was compared against an identical model, except for the main effect of focus location. The model comparison revealed that focus location (i.e., focus on the verb, object, or entire predicate) has a significant effect on the duration of the stressed vowel ($\chi^2(1) = 88.60$, $p< 0.001$). That is, when only the verb or object was focused by itself, the stressed vowel in the verb or object (whichever

was focused) showed a significantly longer duration than the other. However, when the entire predicate is focused, the duration seems to be equally distributed on the two stressed vowels in the predicate (verb and object). A second model was identical to the last full model, except it did not include the interaction term between item location and focus location was compared against the full model. The comparison revealed that the interaction between focus location and item location was significant ($\chi^2(1) = 15.07$, $p< 0.001$). A full model was also compared with a model that did not include focus type. There was no significant main effect of focus type in the data dealing with only words in predicate position ($\chi^2(1) =0.01$, $p= 0.97$), which means there was no significant difference between narrow and contrastive focus.

These differences are illustrated in Figure 2 below. The figure below illustrates that when the verb is under focus (either narrow or contrastive), it shows a longer vowel duration, and the object has a shorter vowel duration. When the object is under focus, it has a longer vowel duration, and the verb has a shorter vowel duration. However, when the entire predicate (verb phrase) is under focus, the duration increase associated with focus seems to be distributed between the two elements of the verb phrase (verb and object).

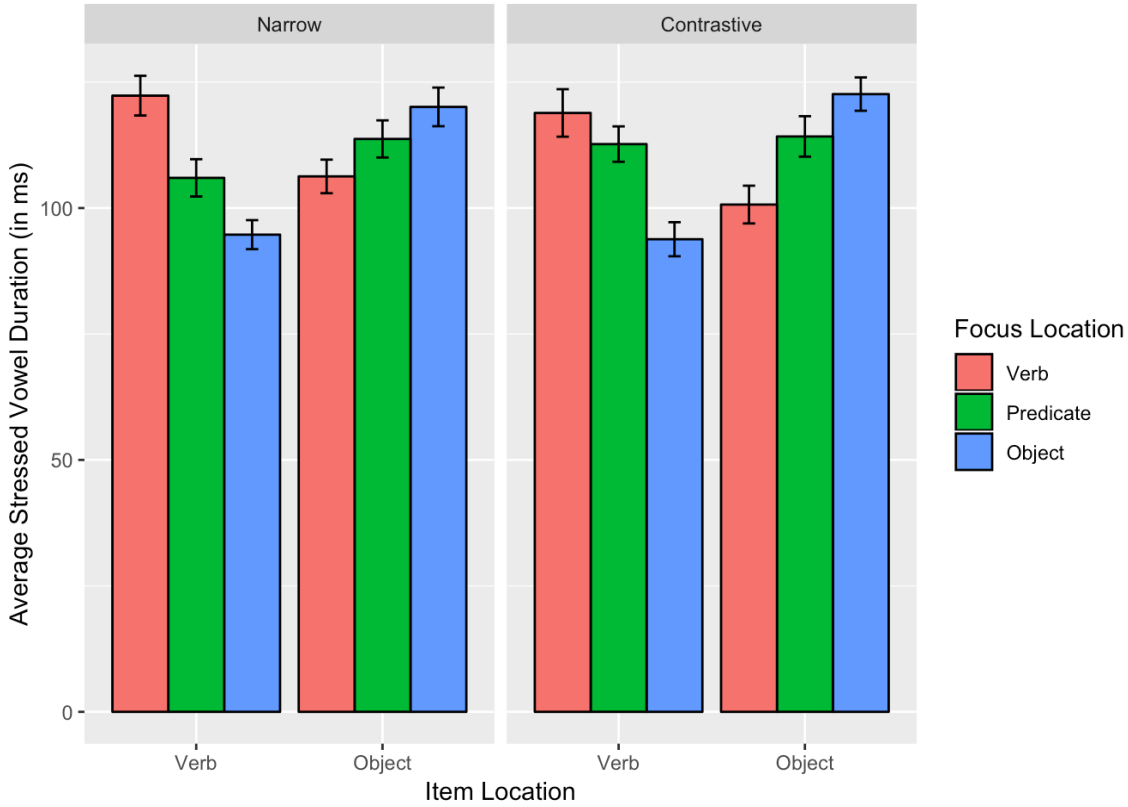

Figure 2 Predicate average stressed vowel duration (ms.) (error bars represent 95% confidence interval).

To summarize, as expected, the duration of stressed vowels was affected by focus status, focus type (but not between narrow and contrastive), and item location. Stressed vowels in words under narrow and contrastive focus were longer than those in words under broad focus across all sentence positions. Conversely, the duration of unfocused (or given) words was significantly shorter than their broad focus counterparts. Additionally, when the predicate was focused, the effect of focus on the duration of the verb and object seemed to be equally distributed between the items, unlike when only one item is focused. Taken together, the findings of this section indicate that focus status affects the

duration of stressed vowels: focused items have longer vowels than unfocused items. Focus type also affects the duration of the stressed vowel: this is particularly relevant in the difference between narrow focus and contrastive focus on the one hand and broad focus on the other.

3.2.2 Intensity

This section reports on the results from the maximum intensity (measured in dB) of the stressed vowel in each target word. Table 8 reports the average maximum intensity results arranged by focus condition, focus type, and item location and provides the standard deviation for each measure. These results are also visualized in Figure 3 below.

Table 8
Target words' average stressed vowel maximum intensity (dB)

Focus Type	Item Location	Focus Status			
		Focused		Unfocused	
		M	SD	M	SD
Broad	*Subject*	66	6.34	NA	NA
	Verb	64.9	7.56	NA	NA
	Object	63.6	7.5	NA	NA
Narrow	*Subject*	67.8	6.68	64.3	5.79
	Verb	66.8	6.85	62.6	6.61
	Object	65	7.5	58.6	6.67
Contrastive	*Subject*	68.3	6.79	64.2	6.09
	Verb	67	7	61.9	7.06
	Object	64.5	7.91	57.6	8.49

Figure 3 shows the maximum intensity for stressed vowels in the target words in dB arranged by focus type (broad, narrow, contrastive), item location (subject, verb, object),

and focus status (focused, unfocused). As can be seen from the figure, overall focused items have higher maximum intensity (M = 65.9, SD = 7.29) than unfocused counterparts (M = 61.9, SD = 7.17). Additionally, items under narrow (M = 66.3, SD = 7.14) and contrastive focus (M = 66.2, SD = 7.46) have higher intensity than items under broad focus (M = 64.8, SD = 7.17).

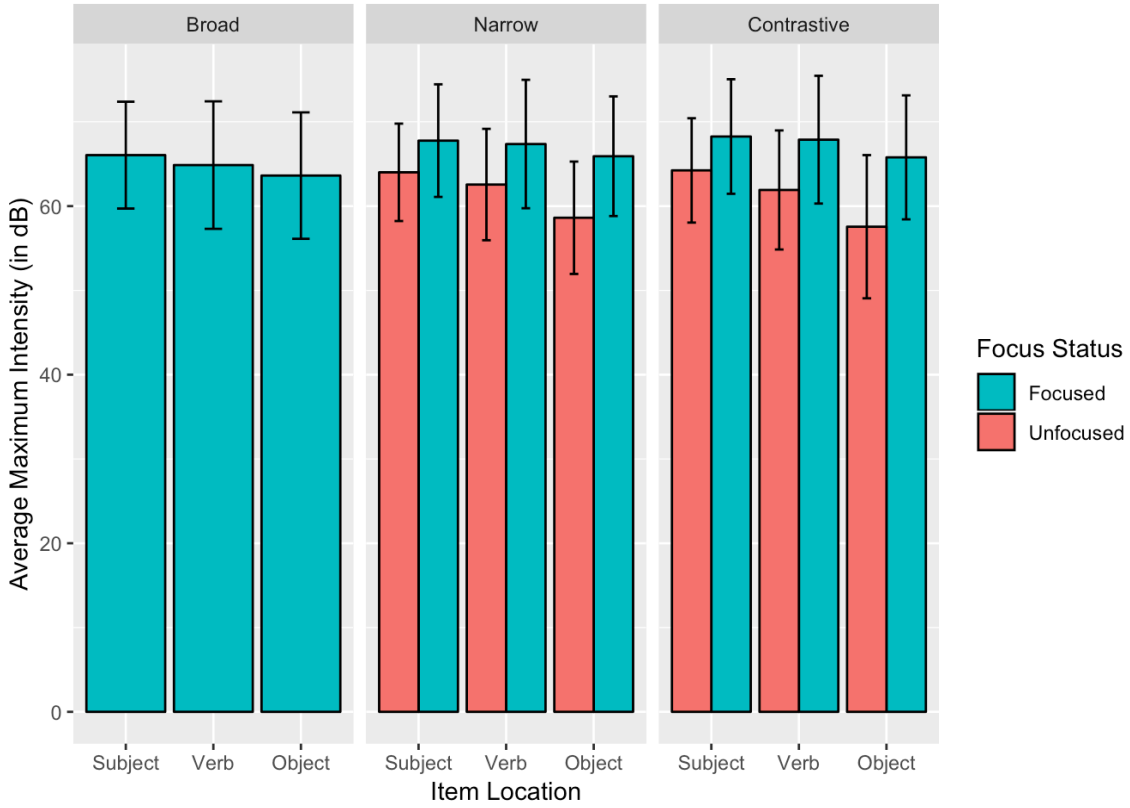

Figure 3 Target words' average stressed vowel maximum intensity (dB) (error bars represent 95% confidence interval).

To test these patterns of maximum intensity, a series of mixed-effects regression models were constructed in R (R Core Team, 2013) using the lmer() function of the lme4

package (Bates et. al., 2013). The independent variables were focus type (broad, narrow, contrastive) focus status (focused, unfocused), and item location (subject, verb, object). The model with the maximal random effects structure failed to converge; therefore, a forward best-path method was used to determine which random slopes to include (Barr, Levy, Scheepers & Tily, 2013). Only focus status met the inclusion criterion (alph = .2), and thus, was included as a random slope with random intercepts for participants and items. The results from the best mixed-effects model are shown in Table 9 with parameter estimate β statistics, standard error, t value, and p-value for the fixed effects.

Table 9
Coefficients of the best linear mixed-effects model of maximum intensity (dB) (N=1296).

	Model Summary			
	β	SE	*t*	*Pr(>|t|)*
Fixed effects				
(Intercept)	63.75	2.27	28.07	**<0.001**
Focus Type (Contrastive)	0.29	0.20	1.45	0.14
Focus Type (Broad)	-1.70	0.31	-5.41	**<0.001**
Focus Status (Focused)	5.03	0.87	5.81	**<0.001**
Item Location (Verb)	3.08	0.23	13.66	**<0.001**
Item Location (Subject)	3.12	0.20	15.45	**<0.001**
Contrastive * Focused	-0.62	0.41	-1.53	0.13
Contrastive * Verb	-0.56	0.47	-1.19	0.23
Broad * Verb	-0.88	0.75	-1.18	0.24
Contrastive * Subject	-0.73	0.45	-1.64	0.10
Broad * Subject	-0.39	0.69	-0.57	0.57
Focused * Verb	-2.03	0.47	-4.29	**<0.001**
Focused * Subject	-1.86	0.45	-4.11	**<0.001**
Contrastive * Focused * Verb	-0.30	0.94	-0.32	0.75
Contrastive * Focused * Subject	0.09	0.90	0.10	0.92
	s^2			
Random effects				
Participant (Intercept)	40.96			
Focus Status (Slope)	5.67			
Item (Intercept)	0.16			

Model comparisons were then conducted to test the significance of the independent factors and their interactions. That is, the full model described above was compared against the model without the effect in question. To this end, models with different predictors were computed and compared with likelihood ratio tests carried out by the ANOVA function, which used log-likelihood as a measure of goodness of fit (cf. Baayen, 2008). Three models were constructed and compared to the full model to test the main effects. The first model was identical to the full model, except it did not include

focus type as a main effect. This comparison revealed a significant main effect of focus type ($\chi^2(1) = 39.27$, $p < 0.001$). This is because the stressed vowels in words under broad focus condition had significantly lower maximum intensity than those under narrow and contrastive focus, as can be seen in the coefficient table above (Table 9). The second model was identical to the full model, except it did not include focus status as a main effect. The comparison revealed that focus status significantly contributed to the model fit ($\chi^2(1) = 52.01$, $p < 0.001$), reflecting that focused items show higher intensity than unfocused counterparts. The third model was identical to the full model, except it did not include the main effect of item location. This comparison revealed a significant main effect of item location ($\chi^2(1) = 425.79$, $p < 0.001$), reflecting a significant downward trend as item location moves from subject to object. This finding shows that words in the subject position generally had higher maximum intensity than words in the verb position, which in turn had higher intensity than words in the object position.

Additionally, one model was constructed to test the interaction between focus status and item location. The model testing the interaction was identical to the full model, except it did not include the interaction between the focus status and item location. This comparison revealed a significant interaction between focus status and item location ($\chi^2(1) = 35.10$, $p < 0.001$), reflecting the significantly decreasing intensity as item location moves from the subject position to the object position in focused and unfocused status, which can be readily observed from Figure 3 above.

Follow-up comparisons were also conducted using contrast coding to further confirm the differences between the three focus types, with the chief goal of exploring

whether there are statistically significant differences between narrow and contrastive focus. First, in order to test whether there is a statistically significant difference between narrow and contrastive focus, broad focus was excluded from the dataset, then compared the full model described above against a model without focus type. This model comparison revealed no significant difference in terms of average maximum intensity between narrow focus and contrastive focus ($\chi^2(1) = 8.95$, $p = 0.11$). Second, in order to compare the differences between broad focus and focused words under narrow and contrastive focus, words that were coded as focused in the focus status variable were included in the dataset. Then, the full model was compared against the model without the focus type variable. The comparison revealed that focus type significantly improved model fit ($\chi^2(1) = 57.03$, $p< 0.001$), indicating that narrowly and contrastively focused words had higher maximum intensity than broadly focused words.

Additionally, another model was constructed to test whether the maximum intensity of the stressed vowels in words under broad focus was different from that of unfocused words. Accordingly, a subset of the data was created that only included the maximum intensity for broad focus words (whose focus status is focused) and the maximum intensity for the unfocused words from both narrow and contrastive focus (whose focus status is unfocused). Two models were then compared where one model had focus status against one without focus status. The model comparison revealed that focus status significantly affected maximum intensity ($\chi^2(1) = 129.87$, $p < 0.01$), indicating that broad focused words had higher maximum intensity than unfocused words. These findings show that there is no statistically significant difference between

narrow focus and contrastive focus in terms of maximum intensity and that broad focus is in an intermediary level between narrowly and contrastively focused items on the one hand and unfocused counterparts on the other.

Further analysis was performed on words in the verb and object position to determine whether focus location (i.e., focus is on the verb, the object, or the entire predicate) has a consequence on the maximum intensity of the stressed vowels on these items (Figure 4). Accordingly, a subset of the data was created that only included narrow and contrastive focus for the focus type and verb, object, and predicate for the focus location variable and verb and object for the item location. That's, broad focus sentences and words in the subject position were excluded. A new full model for this dataset was then created where item location, focus location and focus type, as well as interaction between item location and focus location, were entered as independent variables with participant and sentence are random effects. This new full model was compared against an identical model, except for the main effect of focus location.

This analyses revealed that focus location had a significant effect on the average maximum intensity of stressed vowels in focused items ($\chi^2(1) = 219.49$, $p < 0.001$), indicating that when either the verb or object is focused only by itself it had the highest maximum intensity and when the entire predicate is focused, the maximum intensity of both words in the predicate show relatively higher intensity than the unfocused counterparts. A second model was identical to the new full model, except it did not include the interaction term between item location and focus location was compared against the full model.

The comparison revealed that the interaction was significant ($\chi^2(1) = 19.50$, $p < 0.001$). This shows that when either the verb or the object is focused by itself, it has higher intensity than the other item in the predicate (because it is the only focused item in the predicate). This is predicted, given the difference between focused and unfocused items. However, when the entire predicate is focused, the stressed vowels in both elements (verb and object) show increased intensity compared to the unfocused counterparts, though the verb seems to have greater intensity, presumably because declarative sentences are observed to have a downward trajectory for intensity.

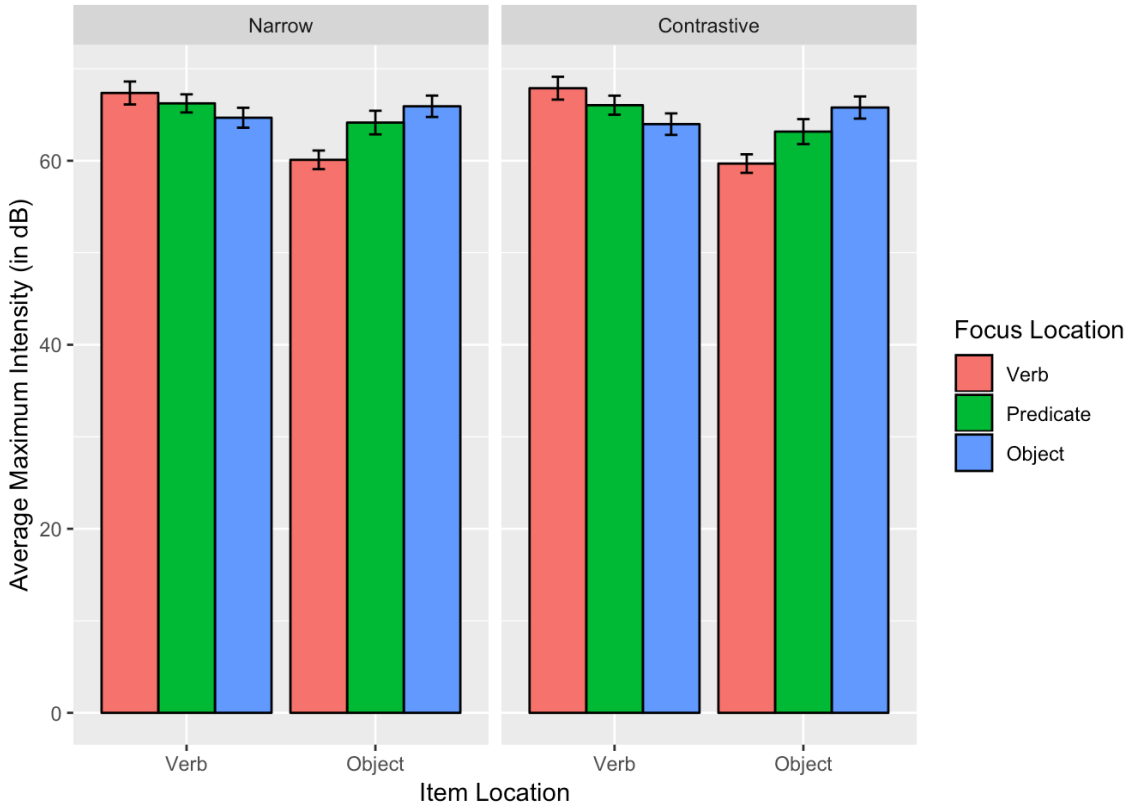

Figure 4 Predicate average stressed vowel maximum intensity (in dB) (error bars represent 95% confidence interval).

To summarize, this section demonstrated that focus status affected the average maximum intensity of the stressed vowels in the target words. Expectedly, focus is correlated with higher intensity. Although no significant differences were found between narrow focus and contrastive focus, broad focus was different from both narrow focus and contrastive focus, in that it had lower maximum intensity. Both focused verbs and focused objects were different from focused predicates, reflecting the expected pattern associated with the size of the focus domain. If both elements in the predicate are under the focus domain, they both show increased intensity compared to their unfocused counterparts, which in turn is lower than the verb's maximum intensity in the verb only focus and object's maximum intensity in the object only focus. Finally, intensity generally followed a downward trajectory as item location moves from subject to object.

3.2.3 Maximum F0

This section reports on the findings from maximum F0 measurements (in semitones). Table 10 below provides a summary of the average maximum F0 of the stressed vowel divided by focus status, focus type, and item location for both females and males. The table also provides the standard deviation for each measure.

Table 10
Target word's average stressed vowel maximum F0 (semitones)

Focus Type	Item Location	Focus Status							
		Focused				Unfocused			
		Female		Male		Female		Male	
		M	SD	M	SD	M	SD	M	SD
Broad	*Subject*	94.4	1.25	90.8	1.73	NA	NA	NA	NA
	Verb	93.6	1.41	90	3.25	NA	NA	NA	NA
	Object	92.5	1.27	88.3	3.48	NA	NA	NA	NA
Narrow	*Subject*	94.8	1.39	92.1	2.39	94	1.4	89.6	1.96
	Verb	94.1	1.62	91.7	2.95	92.7	1.84	87.8	3.26
	Object	92.6	1.43	91.1	4.17	89.9	3.52	85.4	2.81
Contrastive	*Subject*	94.5	1.22	91.3	2.77	93.5	1.37	87.9	2.66
	Verb	93.7	1.3	90.6	3.6	92	2.53	86.8	3.22
	Object	91.9	2.46	90.4	3.82	90.4	2.75	85	2.64

The data in Table 10 are illustrated in Figure 5 below. The figure shows that the average maximum F0 of the stressed vowel in the target words in semitones arranged by focus type (broad, narrow, contrastive), item location (subject, verb, object), focus status (focused, unfocused) and gender of the speaker (male, female). As can be seen from the figure, focused items have higher maximum F0 (M = 92.1, SD = 3.07) than unfocused counterparts (M = 89.8, SD = 3.85). Additionally, items under narrow (M = 92.6, SD = 2.90) and contrastive focus (M = 92, SD = 3.16) have higher maximum F0 than items under broad focus (M = 91.5, SD = 3.08). These patterns seem to be consistent in the productions of both genders. However, overall, female speakers produced higher maximum F0 (M = 92.8, SD = 2.40) than male speakers (M= 89, SD = 3.71).

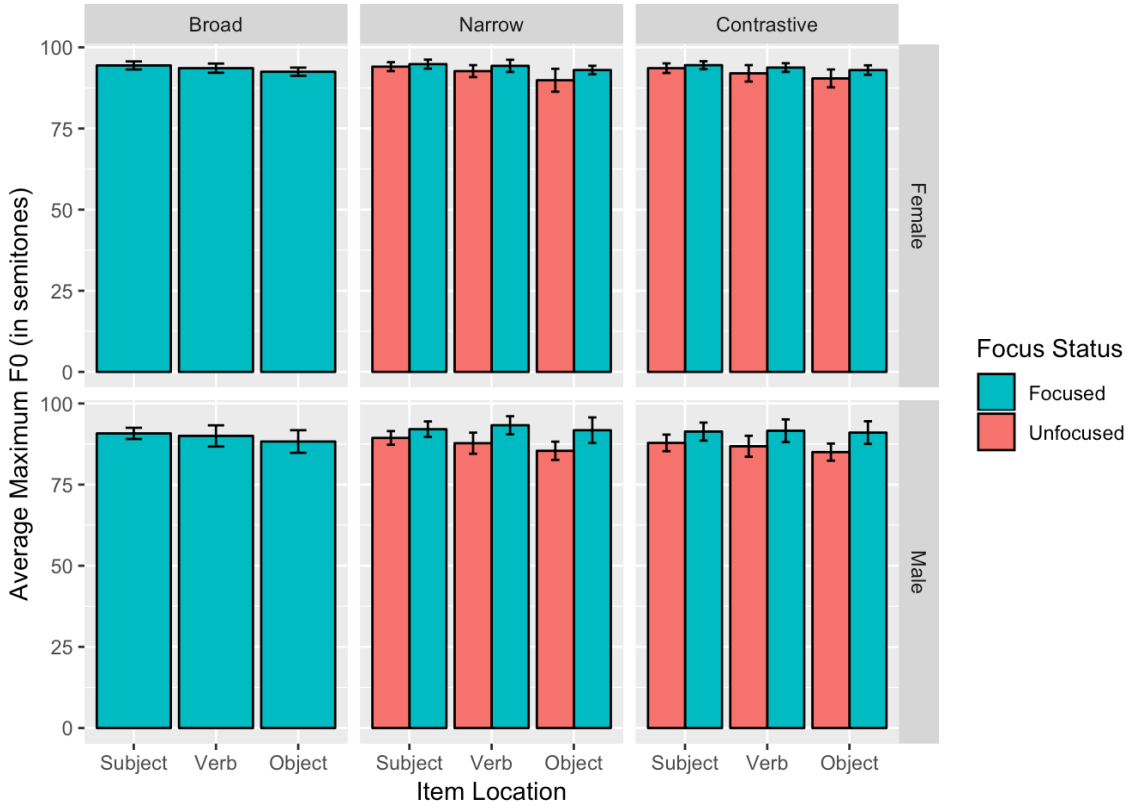

Figure 5 Target words' average stressed vowel maximum F0 (in semitones) (error bars represent 95% confidence interval).

To test these patterns, a series of mixed-effects regression models were constructed in R (R Core Team, 2013) using the lmer() function of the lme4 package (Bates, Maechler & Bolker, 2013). The independent variables were focus type (broad, narrow, contrastive) focus status (focused, unfocused) item location (subject, verb, object) and gender of the speaker (male, female). The model with the maximal random effects structure failed to converge; therefore, a forward best-path method was used to determine which random slopes to include (Barr, Levy, Scheepers & Tily, 2013). Only focus status met the inclusion criterion (alph = .2), and thus, was included as a random

slope with random intercepts for participants and items. The results from the best mixed-effects model are shown in Table 11 with parameter estimate β statistics, standard error, t value, and *p*-value for the fixed effects.

Table 11
Coefficients of the best linear mixed-effects model of maximum F0 (semitones) (N=1296).

	Model Summary			
	β	SE	*t*	*Pr(>|t|)*
Fixed effects				
(Intercept)	90.85	0.70	130.02	**<0.001**
Focus Type (Contrastive)	0.64	0.11	5.63	**<0.001**
Focus Type (Broad)	-0.80	0.18	-4.50	**<0.001**
Focus Status (Focused)	2.81	0.39	7.28	**<0.001**
Item Location (Verb)	1.65	0.13	12.99	**<0.001**
Item Location (Subject)	1.91	0.11	16.76	**<0.001**
Gender (Male)	-3.91	1.38	-2.84	**<0.05**
Contrastive * Focused	0.07	0.23	0.32	0.746
Contrastive * Verb	0.50	0.27	1.86	0.063
Broad * Verb	0.35	0.42	0.84	0.403
Contrastive * Subject	0.27	0.25	1.05	0.292
Broad * Subject	0.32	0.39	0.83	0.406
Focused * Verb	-1.07	0.27	-4.02	**<0.001**
Focused * Subject	-1.32	0.26	-5.19	**<0.001**
Contrastive * Male	0.59	0.23	2.58	**<0.01**
Broad * Male	-1.40	0.36	-3.95	**<0.001**
Focused * Male	2.59	0.77	3.35	**<0.05**
Verb * Male	-0.43	0.25	-1.70	0.089
Subject * Male	-0.29	0.23	-1.25	0.211
Contrastive * Focused * Verb	-0.82	0.53	-1.54	0.124
Contrastive * Focused * Subject	-0.92	0.51	-1.81	0.071
	s^2			
Random effects				
Participant (Intercept)	3.78			
Focus Status (Slope)	1.09			
Item (Intercept)	0.07			

To test the effects of the independent factors and their interactions, a series of model comparisons were conducted where the full model described above was compared against the model without the effect in question. To this end, models with different predictors were computed and compared with likelihood ratio tests carried out by the ANOVA function, which used log-likelihood as a measure of goodness of fit (cf. Baayen, 2008). Fours models were constructed and compared against the full model to test the main effects. The first model was identical to the full model, except it did not include focus type as a main effect. The model comparison revealed a significant main effect of focus type ($\chi^2(1) = 92.41$, $p < 0.001$), indicating that the maximum F0 is different between the three focus types. The second model was identical to the full model, except it did not include focus status as a main effect. The comparison revealed a significant main effect of focus status ($\chi^2(1) = 63.14$, $p < 0.001$), reflecting that focused items show higher maximum F0 than unfocused counterparts. The third model was identical to the full model, except it did not include item location as a main effect. There was a significant main effect of item location ($\chi^2(1) = 453.45$, $p < 0.001$), reflecting a downward trend as item location moves from subject to object. The fourth model was identical to the full model, except it did not include gender as a main effect. The model comparison revealed a significant main effect of gender ($\chi^2(1) = 46.33$, $p < 0.001$), indicating that the female speakers had an overall higher maximum F0 than male speakers.

Additionally, three models testing the interactions were constructed. These models were identical to the full model, except for the interaction term between the two factors. The model comparisons revealed a significant interaction between focus status

and item location ($\chi^2(1) = 42.62$, $p < 0.001$), indicating that a downward trend in the maximum F0 values is different between focused and unfocused words. This is because focus increases the maximum F0 on the focused words and suppresses the maximum F0 on unfocused words. The maximum F0 values are further suppressed for the unfocused words as their position moves from subject to object position. Significant interactions were also found between focus type and gender ($\chi^2(1) = 26.41$, $p < 0.001$) and focus status and gender ($\chi^2(1) = 8.74$, $p < 0.01$), which indicate that the female speakers had higher F0 maximum than male speakers across the different focus types and that the male speakers had a bigger difference in the maximum F0 between focused and unfocused items than female speakers.

Further model comparisons were carried out to explore the differences between the different focus types with the chief goal of exploring whether there are significant differences between narrow and contrastive focus. First, in order to test whether there is a statistically significant difference between narrow and contrastive focus, broad focus was excluded from the dataset, then compared the full model described above against a model without focus type. This model comparison revealed a significant difference in terms of average maximum F0 between narrow focus and contrastive focus ($\chi^2(1) = 16.27$, $p < 0.001$) but gender was not significant ($\chi^2(1) = 3.18$, $p = 0.2$). The unfocused items between narrow and contrastive focus types were compared by including only the observations for unfocused words. The difference in unfocused items between narrow focus and contrastive focus was significant ($\chi^2(1) = 17.70$, $p < 0.001$) with gender also being significant ($\chi^2(1) = 15.33$, $p < 0.001$). That is, a gender difference in narrow and

contrastive focus was found only in unfocused items and that unfocused words in narrow focus sentences were different from unfocused words in contrastive focus utterances. Further, in order to compare the differences between broad focus and focused words under narrow and contrastive focus, words that were as coded as focused in the focus status variable were included in the dataset. Then, the full model was compared against the model without the focus type variable.

The comparison revealed that the difference in terms of maximum F0 between broad focus and focused items under narrow and contrastive focus was a significant one ($\chi^2(1) = 11.60$, $p< 0.001$). Similarly, the differences between broad focus and unfocused items were significant ($\chi^2(1) = 140.08$, $p < 0.001$). These findings are generally similar to those obtained for duration and intensity in that narrowly and contrastively focused items have higher F0 maximum than broadly focused items in the same positions. In turn, broad focus items have higher F0 maximum than unfocused items. However, the main difference is that here contrastively and narrowly focused items showed different F0 maximum.

Further analyses were conducted on the verb and object to determine whether focus location (i.e., focus is on the verb, the object, or the entire predicate) has an effect on the maximum F0 of these items (Figure 6). Accordingly, a subset of the data was created that only included narrow and contrastive focus for the focus type and verb, object, and predicate for the focus location variable and verb and object for the item location. That's, broad focus sentences and words in the subject position were excluded. A new full model for this dataset was created where item location, focus location and

focus type, as well as interaction between item location and focus location, were entered as independent variables with participant and sentence are random effects. This new full model was compared against an identical model, except for the main effect of focus location, which revealed a significant main effect on the maximum F0 of stressed vowels ($\chi^2(1) = 275.05$, $p < 0.001$), indicating that when the predicate is focused both the verb and object have higher F0 maximum. In contrast, when only one item is focused, that item has the highest F0 maximum and the unfocused one has significantly lower F0 maximum. This is expected, given the difference in the focus status of these words. Additionally, there was a significant main effect of focus type ($\chi^2(1) = 50.91$, $p < 0.001$), which indicates that there is a difference between narrow and contrastive focus. This difference comes from the fact that narrow focus had generally higher F0 maximum on the focused items and lower F0 maximum on the unfocused items when compared to contrastive focus.

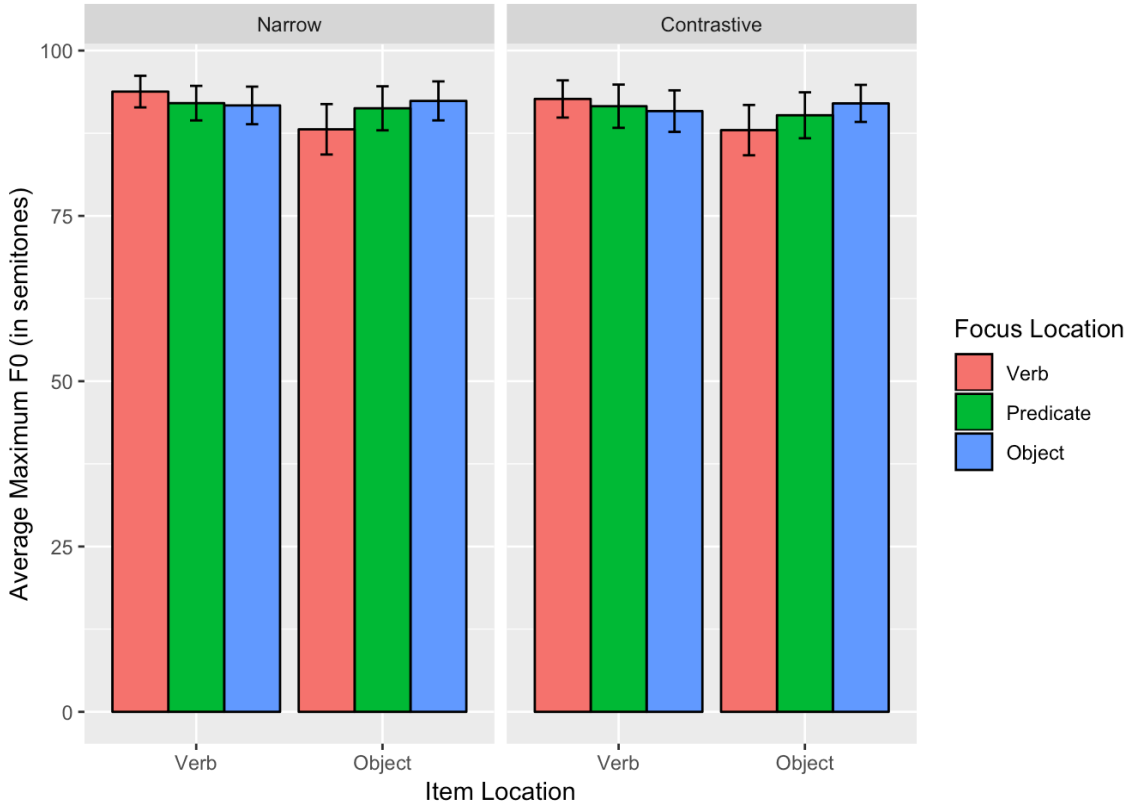

Figure 6 Predicate average stressed vowel maximum F0 (in semitones) (error bars represent 95% confidence interval).

However, the interactions between focus location and item location are more informative in this case. Accordingly, a model without the interaction term between the two factors was compared against the new full model described above. The comparison revealed that the interaction between focus location and item location was significant ($\chi^2(1) = 43.32$, $p < 0.001$). As can be seen in Figure 6, when either the verb or the object was focused, it had higher maximum F0 than the unfocused word; however, when the entire predicate was focused, both the verb and the object had similar F0 maximum, reflecting that both items were equally affected by the focus, which follows the same

pattern observed for duration and intensity. It is worth noting, however, that when the verb is focused, the suppression on the unfocused object seems to be more obvious. Although the focused object shows higher maximum F0, the unfocused verb does not seem to be greatly affected by the unfocused status.

To summarize, as expected, the maximum F0 of stressed vowels was significantly affected by focus status, focus type, item location, and gender. Stressed vowels of words under narrow focus and contrastive focus had higher F0 maximum than words under broad focus, across all sentence positions. Conversely, the maximum F0 in unfocused (or given) words was significantly lower. Additionally, when the predicate was focused, the effect of focus on the maximum F0 of the verb and object seemed to be equally distributed between both items, unlike when only one of them was under focus. Lastly, there were significant differences between narrow and contrastive focus. This difference was not found in the duration and maximum intensity results.

3.2.4 F0 Range

This section reports on the results from the F0 range measurements (in semitones). The average values for the F0 range and standard deviations are summarized in Table 12 split by focus status, focus type, item location for both genders.

Table 12
Target word's average stressed vowel F0 range (semitones

Focus Type	Item Location	Focus Status							
		Focused				Unfocused			
		Female		Male		Female		Male	
		M	SD	M	SD	M	SD	M	SD
Broad	*Subject*	47.8	11.5	47.4	11.9	NA	NA	NA	NA
	Verb	49.2	8.93	46.7	14.1	NA	NA	NA	NA
	Object	46.1	14.2	53.4	12.1	NA	NA	NA	NA
Narrow	*Subject*	52.3	12.6	62.8	8.77	47.5	11.2	46.2	13.9
	Verb	50	12	55.4	12.8	51.9	11.5	41.6	11.5
	Object	52.4	10.9	63.7	10	49.4	15.6	40.6	12.6
Contrastive	*Subject*	53.5	10.5	58.5	10.6	46.2	13.4	38.7	14
	Verb	49	11.2	51.8	17.1	48.4	13.8	39.5	14
	Object	51.7	10.3	59	11.8	50.8	15.5	38.7	14.4

Figure 7 below illustrates the average F0 range values of stressed vowel on the target words by the three focus types (broad, narrow, contrastive), item location (subject, verb, object), focus status (focused, unfocused) and gender of the speaker (male, female). As can be seen from the figure, focused items generally have higher F0 range (M = 53.2, SD = 12.9) than unfocused counterparts (M = 44.9, SD = 14.1). Additionally, items under narrow (M = 55.8, SD = 12.4) and contrastive focus (M = 53.5, SD = 12.9) have higher F0 range than items under broad focus (M = 48.4, SD = 12.3). These patterns are consistent in the productions of both genders; however, male speakers seem to have wider ranges than female speakers when producing focused words and smaller ranges when producing unfocused words.

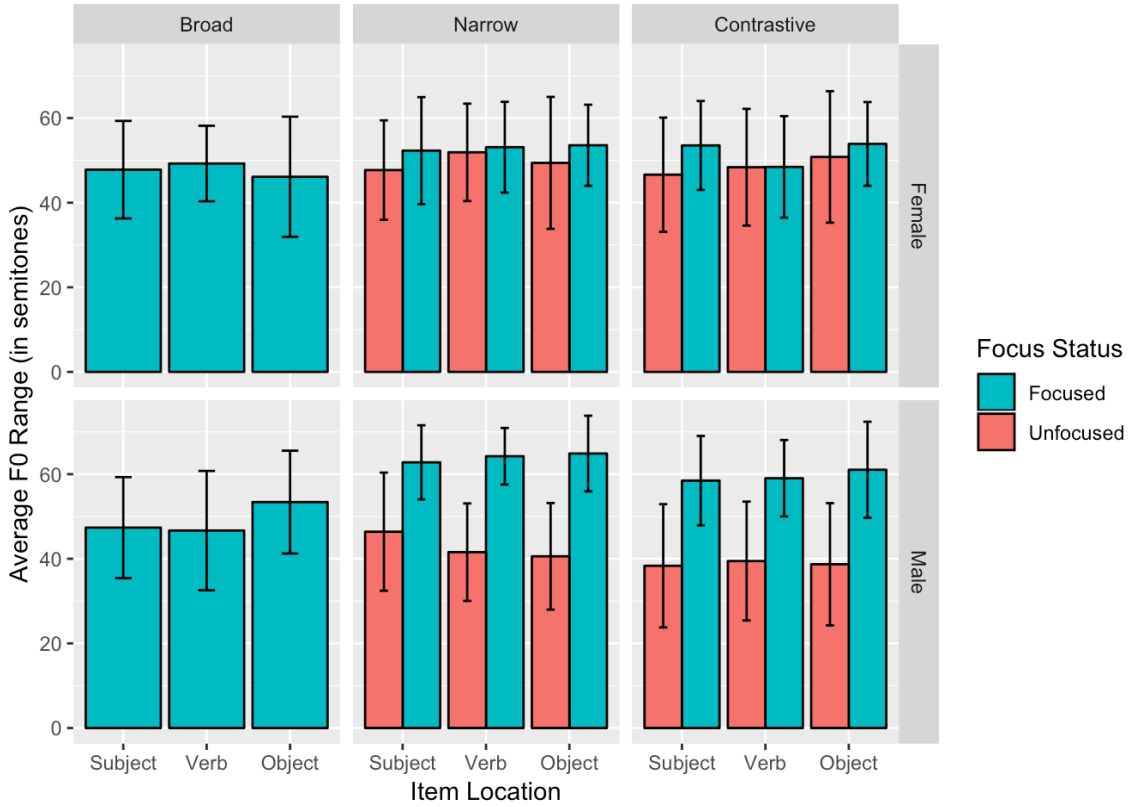

Figure 7 Target words' average stressed vowel F0 range (in semitones) (error bars represent 95% confidence interval).

To test these patterns, a series of mixed-effects regression models were constructed in R (R Core Team, 2013) using the lmer() function of the lme4 package (Bates, Maechler & Bolker, 2013). The independent variables were focus type (broad, narrow, contrastive) focus status (focused, unfocused) item location (subject, verb, object) and gender of the speaker (male, female). The model with the maximal random effects structure failed to converge; therefore, a forward best-path method was used to determine which random slopes to include (Barr, Levy, Scheepers & Tily, 2013). No variable met the inclusion criterion (alph = .2), and thus, no random slope was included

only random intercepts for participants and items. The results from the best mixed-effects model are shown in Table 13 with parameter estimate *β* statistics, standard error, t value, and *p*-value for the fixed effects.

Table 13
Coefficients of the best linear mixed-effects model of F0 range (in semitones) (N= 1296).

	Model Summary			
	β	SE	*t*	*Pr(>|t|)*
Fixed effects				
(Intercept)	49.07	1.56	31.53	**<0.001**
Focus Type (Contrastive)	2.32	0.76	3.07	**<0.01**
Focus Type (Broad)	-6.58	1.18	-5.58	**<0.001**
Focus Status (Focused)	10.06	0.76	13.22	**<0.001**
Item Location (Verb)	-1.86	0.84	-2.20	**<0.05**
Item Location (Subject)	0.60	0.76	0.79	0.43
Gender (Male)	-1.46	2.74	-0.53	0.61
Contrastive * Focused	-0.32	1.52	-0.21	0.84
Contrastive * Verb	1.14	1.77	0.65	0.52
Broad * Verb	3.31	2.80	1.19	0.24
Contrastive * Subject	1.04	1.67	0.62	0.53
Broad * Subject	-3.91	2.58	-1.52	0.13
Focused * Verb	-5.57	1.77	-3.15	**<0.01**
Focused * Subject	3.08	1.69	1.82	0.07
Contrastive * Male	3.36	1.51	2.22	**<0.05**
Broad * Male	-5.63	2.36	-2.39	**<0.05**
Focused * Male	15.20	1.52	9.98	**<0.001**
Verb * Male	-2.60	1.69	-1.54	0.12
Subject * Male	3.01	1.51	1.99	**<0.05**
Contrastive * Focused * Verb	-2.96	3.54	-0.84	0.40
Contrastive * Focused * Subject	-3.79	3.39	-1.12	0.26
	s^2			
Random effects				
Participant (Intercept)	49.07			
Item (Intercept)	2.32			

To test the main effects and the interactions between them, model comparisons, where the full model was compared against a model without the factor in question. To test the main effects, four models were constructed that were identical to the full model, except that they did not include the main effect of focus type, focus status, gender and item location, respectively. Accordingly, the model comparison analyses revealed a significant main effect of focus type ($\chi^2(1) = 59.10$, $p<0.001$), reflecting differences between broad focus, narrow and contrastive focus. This is because narrow and contrastive focus words generally had wider F0 range than broad focus and that narrow focus had wider F0 range than contrastive focus. A significant main effect of focus status ($\chi^2(1) = 262.37$, $p<0.001$) was also found, which reflects that focused items show higher F0 range than unfocused counterparts. The analyses also revealed a significant main effect of item location ($\chi^2(1) = 31.42$, $p<0.001$), reflecting differences in the ranges across the subject, verb and object positions. A significant main effect of gender ($\chi^2(1) = 40.73$, $p<0.001$) was also found, reflecting that the male speakers had wider F0 ranges than female speakers.

Additionally, models testing the interactions between focus status and item location, as well as between focus type and gender, and focus status and gender, respectively, were compared. In these comparisons, models lacking the interaction term between these factors, respectively, were compared against the full model described above. The comparisons revealed a significant interaction between focus status and item location ($\chi^2(1) = 13.28$, $p<0.01$), reflecting that the difference between focused and unfocused words is different across the different item locations. Significant interactions

were also found between focus type and gender ($\chi^2(1) = 12.21$, $p<0.01$), focus status and gender ($\chi^2(1) = 100.51$, $p<0.001$), and item location and gender ($\chi^2(1) = 8.59$, $p<0.05$), indicating the predicted gender difference between male speakers and female speakers across different focus type and focus status. Specifically, male speakers had wider ranges on focused words and smaller ranges on unfocused words whereas female speakers had smaller F0 ranges on focused words than male speakers and the difference in terms of F0 range between focused and unfocused productions by the female speakers is smaller than that of men across the different focus types and item locations.

Follow-up model comparisons were also conducted using contrast coding to confirm the differences between focus types. The comparisons were based on the full model specified above. However, for testing whether there is a difference between narrow and contrastive focus, broad focus data were excluded. Accordingly, the full model was compared with an identical model, except that it did not specify the main effect of focus type. These analyses revealed that the difference between narrow focus and contrastive focus was significant ($\chi^2(1) = 5.03$, $p < 0.05$), reflecting the fact that narrow focus productions had generally wider F0 ranges than contrastive focus.

To test if there is a difference between broad focus and the other two focus types, the full model was compared against a model without focus type. In this comparison, words that were marked as focused for narrow and contrastive focus along with broad focus data were included. The comparison revealed that the differences in terms of F0 range between broad focus and focused words under narrow and contrastive focus were significant ($\chi^2(1) = 70.19$, $p< 0.001$). Similarly, the difference between broad focus and

unfocused words was also significant ($\chi^2(1) = 14.68$, $p < 0.001$). These findings show that the three focus types had different F0 ranges. Items under narrow and contrastive focus had wider F0 ranges than both broad focus and unfocused words. Broad focus items had intermediary F0 ranges between narrow and contrastive focused and unfocused words and that the male and female speakers had different F0 ranges. Overall, these findings follow a similar pattern as the ones observed for F0 maximum, duration and intensity, except for the significant difference between narrow and contrastive focus, which was only found for F0 data. Narrow focus was associated with wider F0 ranges than contrastive focus in this dataset.

Finally, further analyses were conducted to see if focus location as being on the verb, the object or the entire predicate has a consequence on the F0 range (Figure 8). For this additional comparison, a subset of the data was created that only included narrow and contrastive focus for the focus type factor, verb, object, and predicate for the focus location factor and verb and object for the item location. That is, broad focus sentences and words in the subject position were excluded. A new full model was then created where item location, focus location and focus type, as well as interaction between item location and focus location, were entered as independent variables with participant and sentence are random effects. Then models identical to the new full model, except for the factor in question, were created and compared against the new full model. Accordingly, the comparisons revealed that focus location has a significant effect on the F0 range of stressed vowels ($\chi^2(1) = 125.42$, $p< 0.001$), as did focus type significant ($\chi^2(1) =27.93$, $p<0.001$), reflecting that narrow focus words had wider ranges than contrastive focused

counterparts. The interaction between focus location and item location was also significant ($\chi^2(1) = 39.48$, $p < 0.001$), which indicates that focus location increased the F0 range for verbs differently from objects, and predicates. The F0 ranges for objects, and predicates follow the same pattern in that the object seems to show wider F0 range. As can be seen from Figure 8 below, when the verb or the object was focused, it had wider F0 range. When the entire predicate was focused, the F0 ranges of the verb and object were generally smaller than when these two items were individually focused but larger than when they were unfocused. However, in predicate focus, the object not the verb seems to show most of the change in terms of F0 range.

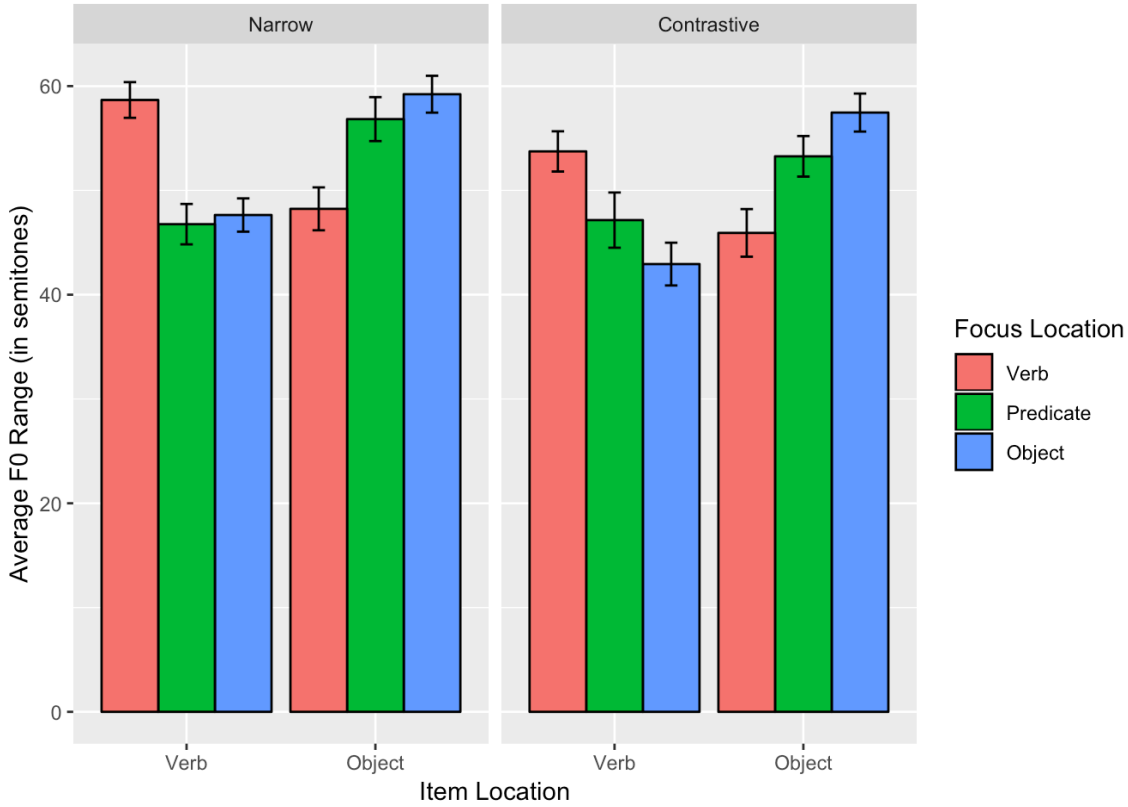

Figure 8 Predicate average stressed vowel F0 range (in semitones) (error bars represent 95% confidence interval).

The findings of this section reveal that there is a significant effect of focus on the F0 range. As expected, focused words had higher F0 ranges than unfocused counterparts. The three focus types were different from each other in terms of F0 ranges and male speakers had higher ranges than female speakers. Predicate focus seems to show a similar pattern as the object focus.

3.3 Discussion

The results from this production reveal that focused words had a significantly longer duration, higher maximum intensity, higher maximum F0 and wider F0 range than

their unfocused counterparts. Further, there are no differences between narrow focus and contrastive focus in terms of duration and intensity. However, F0 analyses reveal significant differences between contrastive focus and narrow focus. Broad focus was shown to be significantly different from narrow and contrastive focus in all of the acoustic measures. Finally, when words in the object and verb position are part of a predicate focus, they both show effects of focus on all of the tested acoustic measures (except for F0 range).

Specifically, in the predicate focus condition, the verb and object were produced in the same intonational phrase, and were both accented, as reflected by the acoustic analysis. Accordingly, the major findings from the production study can be summarized by saying that prosodic focus, as an important aspect of information structure, affected all of the measures analyzed in this production experiment. The findings demonstrate that speakers use specific acoustic cues to distinguish focused words from unfocused ones. That is, speakers use longer duration, higher intensity, higher F0 maximum, and higher F0 range on the stressed vowels of focused words to indicate focus.

These findings show that information structure is marked prosodically by UNA speakers and that duration, intensity and F0 serve as important phonetic correlates of prosodic prominence in this dialect. The findings have also demonstrated that prosodic prominence was achieved by maximizing these phonetic correlates on the target word and minimizing them on the unfocused words (by producing them with shorter duration, lower intensity, and lower F0) so that they are less prominent. Overall, these results are in line with findings from previous studies on other spoken varieties of Arabic in that focus

can be marked prosodically and duration, intensity, and F0 are important to the prosodic prominence (Chahal, 2001; Hellmuth, 2006; Alzaidi, 2014).

In terms of focus type, speakers differentiated broad focus from the other two focus types. Specifically, narrowly and contrastively focused words had significantly longer durations, higher intensity, higher F0 maximum, and wider F0 ranges than words in broad focus. Conversely, unfocused (or given) words had significantly shorter durations, lower intensity, lower F0 maxima and narrower F0 ranges than broad focus. Speakers did not systematically distinguish between narrow and contrastive focus. The distinction was only clear in the F0 results. Specifically, no significant differences were found between narrow and contrastive focus in terms of duration and intensity. Although F0 results yielded significant differences between the two focus types, it is worth noting that these differences consistently followed the same patterns. The fact that speakers did not overwhelmingly differentiate between narrow and contrastive focus could be because of the nature of the task.

In this experiment, speakers heard a pre-recorded question and read the corresponding answer from a computer screen. Such a design is highly controlled, which could have minimized any possible effect of contrast since perhaps the distinction between contrastive and noncontrastive focus requires a more communicative task to appear. It is also possible that speakers were not aware that their productions for contrastive targets were ambiguous between contrastive and noncontrastive. This possibility was actually observed for English in Breen et al. (2010). In their study, they found that only when speakers were made aware of such ambiguity between narrow and

contrastive productions, they differentiate contrastive from noncontrastive utterances as reflected by the acoustic cues they tested in their study.

In terms of the focus breadth, in broad focus, the acoustic correlates were consistently similar across the entire sentence. In narrow and contrastive focus, subject-focus, verb-focus, and object-focus were produced with the longest stressed vowel duration, highest intensity, highest F0 maximum, and widest F0 range compared to their unfocused counterparts. For a wider narrow and contrastive focus domain, as is the case for predicate focus, both the verb and the object generally showed similar measurements of the phonetic features, indicating that speakers were consistently able to indicate the breadth of the focus prosodically. A clear exception to the latter generalization is the F0 range results for predicate focus.

Taken together, UNA speakers were able to mark information structure prosodically in SVO utterances. Speakers differentiated between focused and unfocused words. They prosodically marked the location and breadth of the focus and distinguished broad focus from narrow and contrastive focus, and unfocused words. However, speakers were not able to consistently differentiate between contrastive and noncontrastive utterances. Various acoustic features associated with the prosodic marking of prominence were examined, including duration, intensity, F0 maximum, and F0 range. These acoustic correlates of prosodic prominence were consistently used across the different information structures. These findings confirm that focus is marked prosodically in UNA and that prosodic prominence is achieved by manipulating the acoustic properties, including duration, intensity, and F0. Accordingly, the following chapters are set out to examine

questions related to the perception of prosodic prominence resulting from focus in two perception experiments.

Chapter 4

Perception Experiment I

This chapter reports on a prominence rating experiment examining how well naïve listeners were able to perceive different information structures intended by the speaker from the speech signal alone in UNA. This experiment used data from the production experiment discussed in Chapter 3. In this experiment, listeners took an online survey in which they listened to several UNA SVO sentences produced by two native UNA speakers and rated the degree of prominence of each word on a 5-point rating scale. The sentences represented nine different information structures.

4.1 Method

4.1.1 Material

The material used in this prominence rating task came from the production experiment discussed in Chapter 3. In the production experiment, participants heard pre-recorded questions and read corresponding answers from a computer screen in UNA. For each of the six target sentences, nine set-up questions were asked to elicit different conditions, which resulted in 54 unique productions per talker.

For this perception experiment, the productions elicited from two talkers were used. The talkers were both 26-year-old males, who spoke UNA as their native language. They also spoke English as an L2 and were students at George Mason University at the time of data collection. The purpose of having two different talkers was to test if using different talkers would yield different results. For that reason, it was also important that each talker is heard and rated by a different group of listeners.

The two talkers were selected based on the judgments of five other UNA native listeners with some linguistics training. In this selection process, the five judges sat together. They were told that together they would listen to 432 short utterances produced in UNA by different speakers and judge whether they sounded natural or not. It was explained to them that naturalness is taken to mean that the productions did not sound like reading. The utterances were completely randomized for the speakers. The judges were given a paper with a list of sentences in the same order they heard and marked each utterance on the paper by circling either "natural" or "unnatural." The judges listened to all the recordings and decided that all the productions sounded natural, as no utterances were marked "unnatural."

The judges were then asked to listen together to the four male speakers from the data and choose two speakers that they thought were very different in the way they spoke. The judges selected two male speakers who they thought were very different from each other in that one of them tended to exaggerate his speech while the other did not. The decision to choose male speakers rather than female speakers was mainly based on cultural considerations.

A series of independent sample t-tests were carried out to confirm that the two talkers were different from each other in terms of the acoustic measures of duration, maximum intensity, F0 maximum, and F0 range on all measured words. Likewise, these results were also found when the focus status of the word was considered. The results confirmed that they were indeed significantly different from each other in all measures, except for duration in which they did not differ significantly. The talker who was judged as exaggerating his speech always had higher intensity, higher F0 maximum, and greater F0 range. The t-test results are summarized in Table 14 below.

Table 14
Independent samples T-Test comparison of two speakers

	statistic	**df**	**p**	**Mean difference**	**SE difference**
V1 duration	1.45	322	0.147	4.27	2.939
Maximum F0	-19.10	322	< .001	-6.06	0.317
F0 range	-6.59	322	< .001	-11.17	1.695
Maximum intensity	-30.67	322	< .001	-14.41	0.470

Sig. < .05

The sound files from the two speakers served as experimental utterances in an online prominence rating experiment implemented in Qualtrics, an online survey tool in which a survey can be shared through an anonymous link. For each of the two speakers, 54 target productions representing different focus conditions were examined (namely six utterances for each of the following conditions: broad

focus, narrow subject focus, narrow verb focus, narrow predicate focus, narrow object focus, contrastive subject focus, contrastive verb focus, contrastive predicate focus, and contrastive object focus.)

For the survey design, the first page had the study information, instruction (see details about instruction in the procedure section), and consent form. Then, two main blocks were one for the speaker who exaggerated his speech and one for the speaker who did not. Within each of these two main blocks, a training block, a block for contrastive utterances, and a block for broad and narrow focus utterances were created, which are called noncontrastive because they were not produced in response to a contrastive question. The training block included three training utterances produced by the speaker in that block but were not from the experimental utterances. In the noncontrastive experimental block, there were 30 experimental utterances (six broad focus, six narrow subject focus, six narrow verb focus, six narrow predicate focus, six narrow object focus), and 30 filler utterances produced by the same speaker. The filler utterances closely resembled the experimental utterances but had different syntactic structures. In the contrastive experimental block, there were 24 experimental utterances (six contrastive subject focus, six contrastive verb focus, six contrastive predicate focus, six contrastive object focus), and 30 filler utterances produced by the same speaker. At the end of the experiment, participants were given a short background questionnaire to fill out.

In the survey flow, upon agreeing to participate, participants were automatically assigned to one of the two talkers. Further, for each talker, the contrastive and

noncontrastive blocks were counterbalanced, and the targets were randomized within each block.

4.1.2 Participants

Participants who reported that 1) UNA is their native dialect, 2) they were born in and are from the Riyadh region in Saudi Arabia, 3) completed the survey entirely, and 4) reported no hearing or speech problems were included in this study. Accordingly, 60 participants met the inclusion criteria, and their results from the online prominence rating experiment are analyzed in this chapter. Thirty listeners (15 males and 15 females with ages that ranged between 18 – 49 years (average age = 26.8 years)) listened to and rated the talker who exaggerated his speech.

Twenty of these 30 listeners reported they speak English as L2 and have an average length of residence in an English-speaking country of 1.9 years. Another 30 listeners (15 males and 15 females with an age range between 18 -39 years (average age = 27.3 years) listened and rated the talker who did not exaggerate his speech. 15 of those 30 listeners reported that they speak English as an L2 and have an average length of residence in an English-speaking country of 1.6 years. The assignment for listeners to each speaker block was specified as a condition in the survey flow and done automatically. The reason for not having the same listener listen to both talkers was because it was necessary to see independently if these two different speakers will elicit different prominence ratings without priming the listeners.

4.1.3 Procedure

An anonymous Qualtrics link for the prominence rating experiment was created and shared with the participants. The prominence rating task was a modified version of Rapid Prosody Transcription (RPT) (Cole et al., 2016) adapted from Bishop (2012). According to Cole & Shattuck-Hufnagel (2016), RPT is described as a simple transcription and annotation method developed to collect perceived prominence and boundary strength ratings from untrained transcribers based on what they hear in the audio file. In typical RPT, prominence and boundary tasks are administered separately. Transcribers are presented with unmarked transcripts of the utterances in the audio files. For prominence rating, for example, transcribers are given minimal instructions to mark or highlight prominent words that they think stand out. Transcribers are told that there is no right or wrong answer and that different listeners may perceive prominence on the same utterance differently. Then, they listen to the entire audio files first and rate prominence in real-time. The transcribers' ratings are used to assign a prominence score (p-score) for each word on its aggregated prominence ratings. Higher p-scores reflect higher inter-transcriber agreement and higher prominence. Zero or near 0 p-scores reflect that the word was not prominent. Boundary ratings follow similar logic. There are, of course, other details that can be modified for RPT to serve different purposes.

In the modified version of the RPT task used in the current experiment, naïve listeners were given minimal instructions and asked to rate every word in the utterance on a 5-point rating scale, instead of only marking the prominent word. Once participants in the current study accessed the link, they read the consent form, which included general

information about the study. Having agreed to participate, participants were automatically assigned to one of the two talkers. Listeners were informed that they would listen to approximately 117 short utterances (3 training utterances, 54 experimental utterances, and 60 filler utterances) in UNA. Their task was to rate how stressed or emphasized each word was using the rating scale provided for each word from 1 – 5, where 1 means not at all stressed and 5 means very stressed. The following is a translation of the instructions they received:

"This study is concerned with how speakers pronounce words in a sentence. In the following, you will listen to approximately 117 short utterances spoken in UNA, and your task is to rate as accurately as possible how stressed/emphasized/important each word sounds relative to other words in the sentence using the sliding scale provided, where 1 means least stress and 5 means most stressed."

Listeners were able to proceed through the survey at their own pace and were allowed to listen to each sound file more than once. They were instructed to wear headphones and be in a quiet room while taking the survey. Before beginning the experiment, listeners were given three practice utterances to familiarize themselves with the process and rating scale. The practice utterances were produced by the same speaker and resembled the structure of the experimental data and were not included in the actual experiment. After the practice session was completed, participants started the experiment and provided their ratings for every word in each sentence. At the end of the survey, a short demographic questionnaire was presented (Appendix B). The experiment lasted

approximately 25 minutes. Figure 9 below shows a screenshot of what participants saw in the rating task.

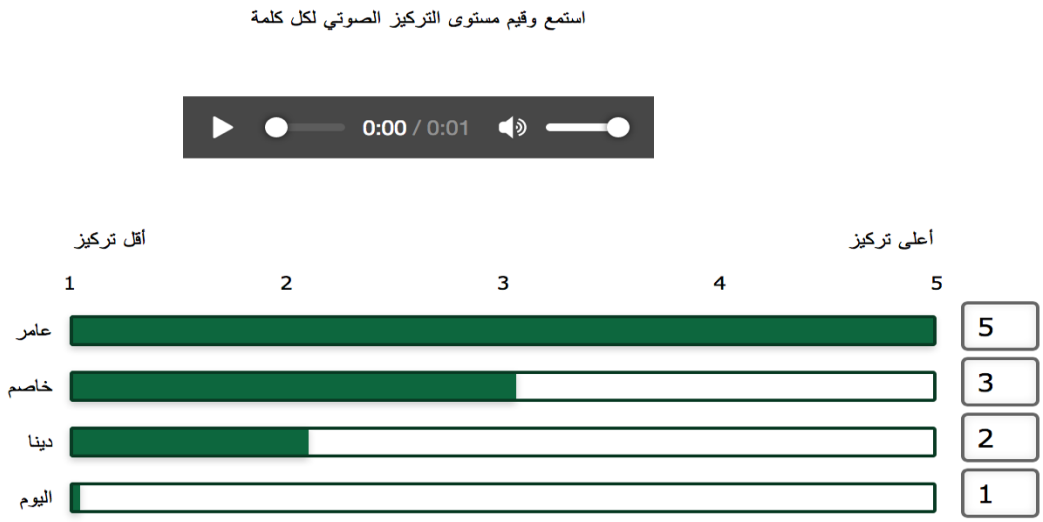

Figure 9 Perception experiment I interface

At the top of the screen, the statement “listen and rate stress/emphasis for each word” appeared above the audio file. Below the radio buttons for the audio file was the scale presented horizontally with the statements “least stress” and “most stress” appeared at the edges along with the scale points from 1- 5. Vertically, on the left side of the screen, each word was written in Arabic orthography in the same order as they were played in the audio files (from the top: subject, verb, object, adverb). Listeners used the mouse to move the sliding scale for each word, and the squares on the right automatically displayed the rating the rater chose on the scale. The numbers were displayed merely for accuracy purposes. The survey could be completed using a computer, or a smartphone, and listeners were not allowed to go back once an utterance was rated.

4.1.4 Data Analysis

The prominence ratings served as the dependent variable in mixed-effects regression models implemented in the program R (Baayen et al., 2008; R Core Team, 2014), using the lmer() function of the lme4 package (Bates, Maechler & Bolker, 2013). Models with different independent (or predictor) variables and random effects were constructed and compared with likelihood ratio tests using the ANOVA function. Variables that significantly improved model fit were retained in the best-fitting model. For the model, the independent variables always included focus type (broad, narrow, contrastive) focus status (focused, unfocused) item location (subject, verb, object). Talker and order did not significantly improve model fit, and thus, were not included in the best-fitting model. The model with maximal random effects structure failed to converge. Therefore, a forward best-path method was used to determine which random slopes to include (Barr, Levy, Scheepers & Tily, 2013). As a result, focus type, focus status and item location (with no interaction between them) met the inclusion criterion (alph = .2), and thus, were included as random slopes with random intercepts for participants and items. The results from the best model are shown for each measure with parameter specific *p*-values obtained by using the Satterthwaite approximation, implemented in the lmerTest package in R (Kuznetsova et al., 2017).

4.2 Results

Table 15 below shows the average prominence ratings, and standard deviations divided by focus type (broad, narrow, contrastive), item location (subject, verb, object), and focus status (focused, unfocused). The results from the table are visualized in Figure 10.

Table 15
Average prominence ratings from perception experiment I.

Focus Type	Item Location	Focus Status			
		Focused		Unfocused	
		M	SD	M	SD
Broad	*Subject*	3.75	1.11	NA	NA
	Verb	3.52	1.03	NA	NA
	Object	3.21	1.09	NA	NA
Narrow	*Subject*	4.59	0.895	3.2	1.11
	Verb	4.06	1.09	3.23	1.06
	Object	3.79	1.23	2.93	1.04
Contrastive	*Subject*	4.56	0.806	3.18	1.05
	Verb	3.83	1.16	3.06	0.958
	Object	3.94	1.2	2.86	0.961

As can be seen from Figure 10, overall focused items received higher prominence ratings (M = 3.84, SD = 1.16) than unfocused counterparts (M = 3.09, SD = 1.04). Additionally, items under narrow (M = 4.06, SD = 1.15) and contrastive focus (M = 4.02, SD = 1.15) have higher prominence ratings than items under broad focus (M = 3.50, SD = 1.10). The figure also shows a downward trend of prominence ratings as item location

moves from subject to object, consistent with the expected prominence trend in the production of declarative utterances.

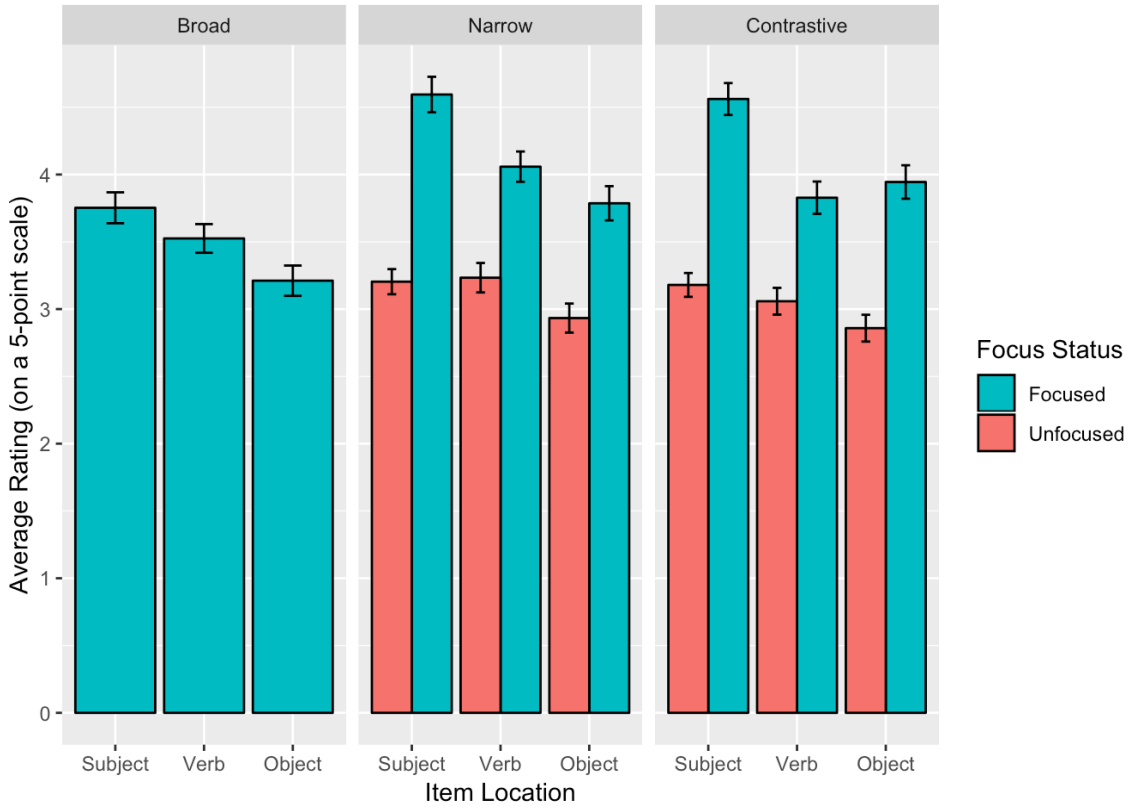

Figure 10 Perception experiment I: average prominence ratings (error bars represent 95% confidence interval).

To test these patterns, a series of mixed-effects regression models were constructed in R (R Core Team, 2013) using the lmer() function of the lme4 package (Bates, Maechler & Bolker, 2013). The independent variables were focus type (broad, narrow, contrastive) focus status (focused, unfocused) and item location (subject, verb, object). The focus type variable represents the focus condition of the sentence, either broad, or narrow, or contrastive. The focus status

variable represents the status of each word in the sentence as being focused or not focused. The item location variable represents whether the word is in the subject, verb, or object position independent from its focus condition. The interactions of these variables give the nine possible conditions described in the method section of Chapter 3 (section 3.1.2). Focus status, focus type, and item location were included as random slopes with random intercepts for participants and items. The results from the best model are shown in Table 16 below, with standard error, *t*-value, and *p*-value.

Table 16
Coefficients of the best linear mixed-effects model of prominence rating-experiment I (N= 5400).

	Model Summary			
	β	SE	*t*	*Pr(>|t|)*
Fixed effects				
(Intercept)	3.39	0.07	48.35	**<0.001**
Focus Status	1.05	0.09	12.31	**<0.001**
Focus Type (contrastive)	0.07	0.05	1.27	0.21
Focus Type (broad)	-0.63	0.06	-10.38	**<0.001**
Item Location (Verb)	0.24	0.05	5.16	**<0.001**
Item Location (Subject)	0.33	0.06	5.50	**<0.001**
Focused * Contrastive	-0.06	0.06	-1.02	0.31
Focused * Verb	-0.17	0.06	-2.69	**<0.01**
Focused * Subject	0.50	0.06	8.20	**<0.001**
Contrastive * Verb	0.24	0.06	3.82	**<0.001**
Broad * Verb	0.24	0.08	3.01	**<0.01**
Contrastive * Subject	-0.05	0.06	-0.85	0.40
Broad * Subject	-0.29	0.08	-3.85	**<0.001**
Contrastive * Focused * Verb	0.29	0.13	2.26	**<0.05**
Contrastive * Focused * Subject	0.10	0.12	0.80	0.42
	s^2			
Random effects				
Participant (Intercept)	0.28			
Focus Status (Slope)	0.39			
Focus Type (Slope)	0.26			
Item Location (Slope)	0.26			
Item (Intercept)	0.01			

To test the significance of these factors and their interactions, model comparison analyses were conducted. In the model comparisons, the full model (the model with the best fit described above) was compared against the model without the effect in question. Accordingly, models with different predictors were computed and compared with likelihood ratio tests carried out by the ANOVA function, which used log-likelihood as a measure of goodness of fit (cf. Baayen,

2008). First, three models were created and compared against the full model to test the main effects. The first model was identical to the full model, except it did not include focus type as a main effect. The comparison revealed a significant main effect of focus type ($\chi^2(1) = 103.01$, $p < 0.001$), reflecting that the prominence ratings of broad focus sentences were different from those of the two other focus types, as can be seen from the coefficient table above. The second model was identical to the full model, except it did not include focus status as a main effect. This comparison also revealed a significant main effect of focus status ($\chi^2(1) = 149.14$, $p < 0.001$), reflecting that focused words received higher prominence ratings than unfocused counterparts. The third model did not include the main effect of item location. The comparison revealed a significant main effect of item location ($\chi^2(1) = 129.1$, $p < 0.001$), reflecting that prominence ratings generally decreased as item location moved from subject to object, as can be seen from Figure 10 above.

Additionally, two models were constructed to test the interactions between the factors in the same way main effects were tested, except this time, the interaction terms for the models were not included and compared against the full model. These comparisons revealed a significant interaction between focus type and item location ($\chi^2(1) = 39.75$, $p < 0.001$), which indicates that words in the subject, verb and object positions received higher prominence ratings under narrow and contrastive focus compared to words in the same position in broad focus utterances. A significant interaction was also found between focus status and item location ($\chi^2(1) = 73.92$, $p <$

0.001), which reflects that item location influences the prominence ratings in focused words more than it does in unfocused ones.

Further follow-up comparisons were conducted using contrast coding to determine which focus types were significant from each other. First, in order to test whether there is a statistically significant difference between narrow and contrastive focus, broad focus was excluded from the dataset, then compared the same full model described above against a model without focus type. The analyses revealed that the difference in terms of prominence ratings between narrow focus and contrastive focus was not significant ($\chi^2(1) = 2.09$, $p = 0.35$). Additionally, in order to compare the differences between broad focus and focused words under narrow and contrastive focus, words that were coded as focused in the focus status variable were included. Then, the full model was compared against the model without the focus type variable. The comparison revealed that the difference between broad focus and focused words under narrow and contrastive focus was significant ($\chi^2(1) = 59.15$, $p< 0.001$). Another comparison was carried out between broad focus and unfocused words in narrow and contrastive sentences, using only the unfocused words from the narrow and contrastive focus types. This comparison revealed a significant difference in terms of prominence ratings between broad focus and unfocused words in narrow and contrastive sentences ($\chi^2(1) = 40.58$, $p < 0.001$).

These comparisons show that narrow and contrastive focus were not rated differently and that broad focus was different for both narrowly and contrastively

focused words, as well as unfocused words. That is, words under broad focus were rated to have prominence that is significantly lower than focused words under narrow and contrastive focus and significantly higher than the unfocused words. This pattern is expected because broad focus represents the neutral unmarked prominence. Therefore, broad focus is expected to have an intermediary perceived prominence level between focused and unfocused words.

Further analyses were conducted on items in the verb and object positions to determine whether focus location on the verb, the object or the entire predicate has a consequence on prominence rating. To achieve this goal, a subset of the data was created that only included ratings for narrow and contrastive focus for the focus type, verb, object, and predicate for the focus location variable and verb and object for the item location. A new full model for this dataset was created, where item location, focus location, and focus type, as well as the interaction between item location and focus location, were entered as independent variables with participant and sentence are random effects. This new full model was compared against identical models, except that they were not specified for the main effect of focus location or interaction between focus location and item location.

The analyses revealed that focus location has a significant main effect on the prominence rating in focused items ($\chi^2(1) = 6.71$, $p< 0.05$). The interaction between focus location and item location was also significant ($\chi^2(1) = 227.47$, $p< 0.001$). These differences are illustrated in Figure 11 below. The figure below illustrates that when the verb is under focus (either narrow or contrastive), it receives a higher prominence rating

than the object. Similarly, when the object is under focus, it receives a higher prominence rating than the verb. This effect is obviously expected, given the difference in focus status. However, when the entire predicate (verb phrase) is under focus, both the verb and object are generally perceived with higher degrees of prominence than their unfocused counterparts. This predicate prominence is not as high as when only one element is focused but is certainly significantly higher than when these items are unfocused. Figure 11 also shows that the object in the predicate focus in contrastive focus was perceived to be more prominent than the verb in the same condition.

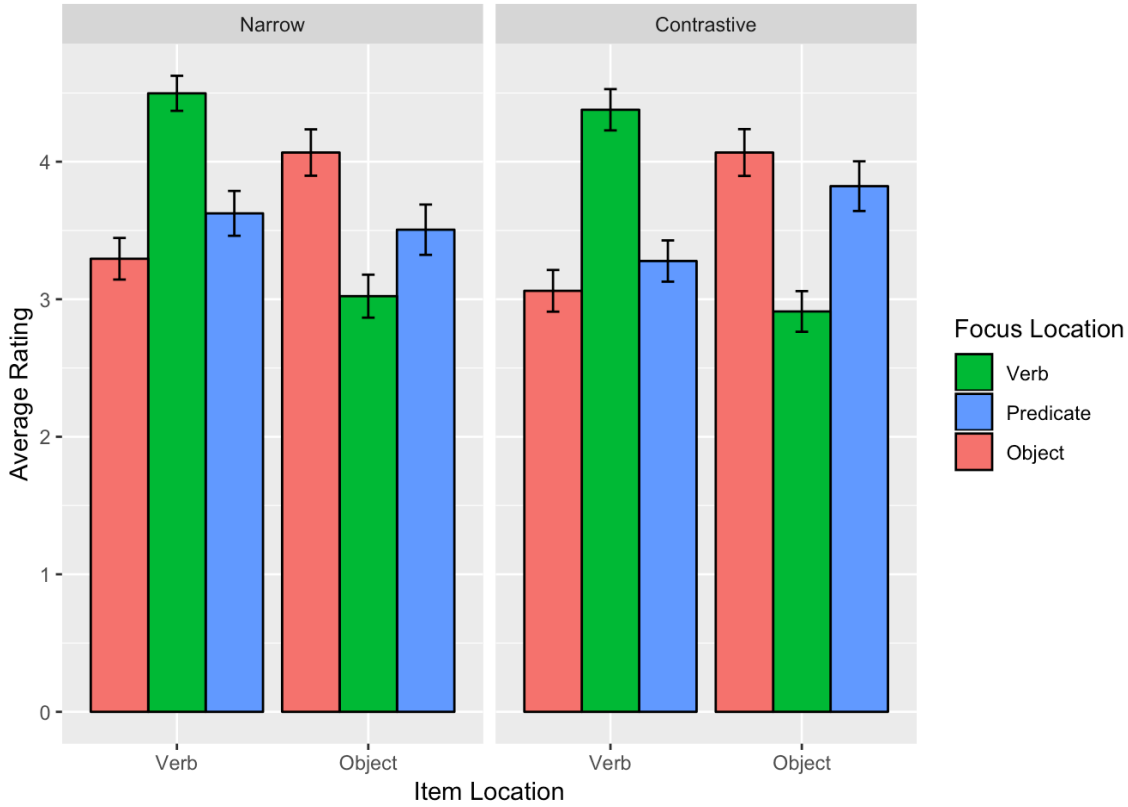

Figure 11 Perception experiment I: predicate average prominence ratings (error bars represent 95% confidence interval).

To summarize, listeners assigned higher prominence ratings to focused words than unfocused words. Listeners did not distinguish between narrow and contrastive focus but distinguished these two focus types from broad focus. Words under broad focus were perceived to be more prominent than unfocused words in the sentences with a narrow or contrastive focus. Additionally, the prominence ratings generally show a perceived downward trajectory as item location moves from subject to object, which is expected in declarative utterances. As for the differences among predicate, verb, and object focus, the results showed that when the entire predicate was focused, listeners

perceived prominence on both the verb and object, as shown by the difference between the perceived ratings for predicate and unfocused counterparts.

Taken together, the findings of this section indicate that listeners were able to successfully retrieve the information structures encoded by the speakers based on the speech signal alone. As reflected by their prominence ratings, listeners were able to distinguish focused from unfocused words and identify the location and size of the focus. They, further, distinguished broad focus from focused words under narrow and contrastive focus. In the production, speakers did not overwhelmingly distinguish between narrow and contrastive focus as significant differences were only found for F0 measures. In this perception experiment, listeners did not significantly distinguish these two focus types.

4.3 Discussion

The production experiment has shown that UNA speakers made distinctions between different information structures prosodically. In the production experiment, speakers manipulated phonetic cues such as duration, intensity, and F0 to mark prosodic prominence. In this perception experiment, listeners were asked to listen and rate prominence on a number of utterances representing different focus conditions. The primary goal was to explore how well naïve listeners are able to decode the information encoded by the speakers in production without any information about the context. That is, will listeners be able to successfully perceive the prominence marked by the speakers from the speech signal alone with no context? It was hypothesized that listeners would

actually be able to perceive such information encoded by the speakers because, in the absence of discourse information, listeners would attend to the information in the acoustic signal.

The perception results from this study mirror the production results such that listeners were successful in distinguishing between broad focus on the one hand and narrow and contrastive focus on the other. Listeners were also highly successful in recognizing different focus locations. These results suggest that UNA listeners were able to exploit the cues used by speakers to signal different information structures to identify prominence.

In this experiment, listeners' ratings of perceived prominence are signal-based. Hence, the findings from this study suggest that focus, as part of information structure, enhances specific acoustic cues in production, and these cues modulate the perception of the information structure, at least in the absence of a context when the listeners have access only to the information in speech signal alone. In the following chapter, another perception experiment is designed to test whether context could influence the perception of prosodic prominence.

Chapter 5

Perception Experiment II

This chapter reports on a prominence rating experiment examining whether naïve UNA listeners' expectations on how different aspects of the information structure are prosodically marked in production can influence their perception of prosodic prominence. In this experiment, listeners took an online survey in which they read questions (facilitating context) and listened to short SVO utterances as produced by a UNA speaker representing answers to the corresponding questions. The auditory stimuli in this experiment used a subset of the data from the production experiment discussed in Chapter 3. The listeners' task was to rate the degree of prominence of each word on a 5-point rating scale, similar to those in the perception experiment 1 in Chapter 4. The sentences represented different information structures, as will be shown in section 5.1.1.

5.1 Method

5.1.1 Material

The material used in this prominence rating task came from the production experiment discussed in Chapter 3. In the production experiment, participants heard pre-recorded questions and read corresponding answers from a computer screen in UNA. For each of the six target sentences, nine set-up questions were

asked to elicit different conditions (see chapter 3 section 3.1.2), which resulted in 54 unique productions per speaker. Only the utterances originally produced for narrow focus (noncontrastive) condition by one male speaker (the speaker who exaggerated his speech from Chapter 4) were used in this experiment.

The purpose of this experiment was to examine if listeners' expectations about how information structure should be realized (according to the context question) affect the perceived prominence. To this end, listeners saw a set-up question (serving as a context) and heard a corresponding answer to that question. The set-up questions and corresponding answers were either context-signal matching or context-signal mismatching. That is, if the question is asking about the subject, and the answer in the audio was originally produced in response to that question, the question-answer pair is considered a "context-signal matching pair." However, if the question is asking about the object, for example, but the answer in the audio file was originally produced as an answer to a question about the subject, the pair is considered a " context-signal mismatching pair."

In the context-signal matching condition, there were 2 target sentences each produced in response to 4 questions eliciting focus in 4 different sentential positions (subject, verb, object, predicate), resulting in 8 context-signal targets. These answers were appropriate to their preceding questions in terms of prosodic organization. To exemplify, if the context question asks about the subject as in *who scolded Sarah yesterday?* then the subject will be prominent in the answer *THAMER scolded Sarah yesterday.* Similarly, if the context question asks about the object as in *Who(m) did*

Thamer scold yesterday? then the object will be prominent in the answer *Thamer scolded SARAH yesterday.* The same is also true for the verb, and predicate focus.

In the context-signal mismatching condition, the answer is not prosodically appropriate to the question. For example, if the question asks about the subject as in *who texted Lina yesterday?* the answer will have prominence on the predicate rather than the subject as in *Sami TEXTED LINA yesterday*.

Since the purpose of the context-signal mismatching experiment was to examine the independent effect of the context on the perception of prosodic prominence, it was critical to hold all the acoustic information constant in this condition. This was achieved by extracting the answer originally produced in response to the predicate focus question (e.g., *what did Sami do yesterday?*) and use it as the answer for all the different set-up questions. That is, although listeners will see different set-up questions serving as contexts (asking about the subject, the verb, the object), the audio file they will hear as the answer for these questions is the same. The predicate focus answers were chosen to serve as answers for the different questions in the context-signal mismatching condition because in the experimental sentences in this condition both the verb and object were always produced in the same intonational phrase and were both equivalently accented with the nuclear accent being on the object and the prenuclear accent being on the verb as reflected by the acoustic analysis. Additionally, words in the subject position in these sentences were not prominent nor did they carry any pitch accent. For the context-signal mismatching condition, 4 target utterances

with predicate focus each was used after 4 set-up questions (1 eliciting focus on the subject, 1 on the verb, 1 on the object, and 1 on the predicate as a control). For example, the same utterance *Sami TEXTED LINA yesterday* was invariably used after the following set-up questions: 1. *Who texted Lina yesterday?* 2. *What did Sami do to Sarah yesterday?* 3. *Who(m) did Sami text yesterday?* and as a control 4. *What did Sami do to Lina yesterday?*

The written set-up questions and the sound files served as experimental material in an online prominence rating experiment implemented in Qualtrics. The 24 experimental question-answer pairs (8 question-answer pairs in the context-signal matching pairs and 16 question-answer pairs in the context-signal mismatching condition) were intermixed with 20 filler question-answer pairs (10 context-signal matching, 10 context-signal mismatching). The fillers closely resembled the experimental utterances but had different syntactic structures. The data were presented in one block.

5.1.2 Participants

Participants who reported that 1) UNA was their native dialect, 2) they were born in and are from the Riyadh region in Saudi Arabia, 3) completed the survey entirely, and 4) did not participate in any of the previous experiments, were included in this study. Accordingly, 36 participants met the inclusion criteria, and their results from the online prominence rating experiment are analyzed in this chapter (16 males and 20 females with an age range between 18 – 36 years and an average age of 24.6 years). All of the listeners reported that they speak English as an L2 and had an average length of residence in an

English-speaking country of 1.8 years. All of the participants reported no hearing of speech problems.

5.1.3 Procedure

An anonymous Qualtrics link for the prominence rating experiment was created and shared with the participants. The prominence rating task was a modified version of Rapid Prosody Transcription (RPT) (Cole, et al 2016) described in Chapter 4.

In this modified version of the RPT task, naïve listeners were given minimal instructions and asked to read the questions and listen to the corresponding answers. Their task was to rate every word in the answer on a 5-point rating scale. Accordingly, once participants in the current study accessed the link, they read the consent form, which included general information about the study. Listeners were informed that they will read a set of approximately 50 questions and listen to approximately 50 short answers to these questions in UNA and their task was to rate how stressed or emphasized each word in the answer part was, using the rating scale provided for each word from 1 – 5, where 1 means not at all stressed and 5 means very stressed. The following is a translation of the instructions they received:

"This study is concerned with how speakers pronounce words in a sentence. In the following, you will read approximately 50 short questions and listen to their corresponding answers spoken in UNA, and your task is to rate as accurately as possible how stressed/emphasized/important each word sounds relative to other words in the

sentence using the sliding scale provided, where 1 means least stressed and 5 means most stressed."

Listeners were able to proceed through the survey at their own pace and were allowed to listen to each sound file more than once. They were instructed to wear headphones and be in a quiet room while taking the survey. Before beginning the experiment, listeners were given three practice utterances to familiarize themselves with the process. These practice utterances were not used in the actual experiment. After the practice session was completed, participants started the experiment and provided prominence rating for every word in the utterances they heard (i.e., listeners rated prominence of the subject, verb, object, and adverb for every utterance they heard) by clicking on the bar. At the end of the survey, a short demographic questionnaire was included. The experiment lasted approximately 20 minutes. Figure 12 below is a screenshot of what participants saw in the rating task.

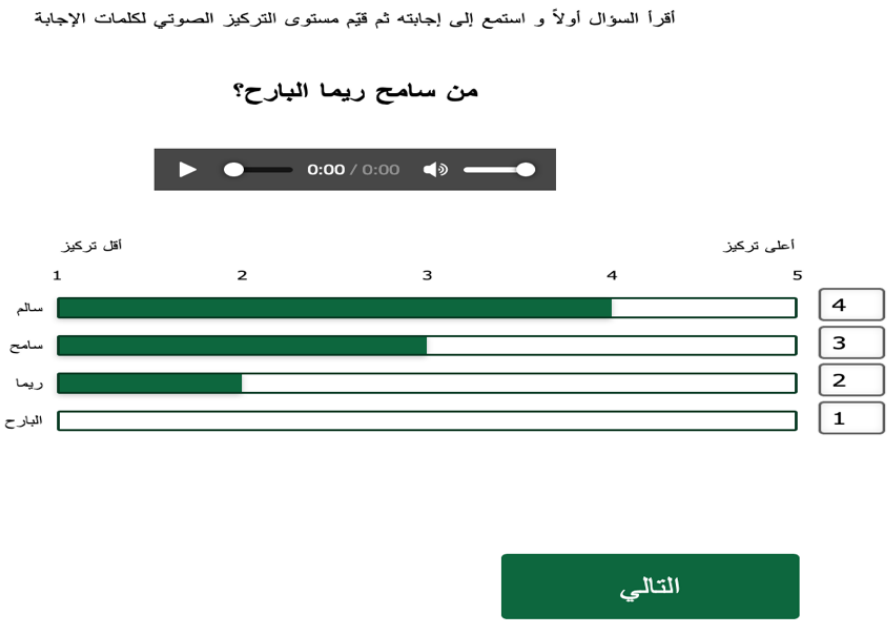

Figure 12 Perception experiment II interface

At the top of the screen, the statement “Read the question first, then listen and rate stress/emphasis for each word in the answer” appeared. Above the audio file, there was the context question written in Arabic. Below the radio buttons for the audio file, there was the scale presented horizontally with the statements “least stress” and “most stress” appearing at the edges along with the scale points from 1- 5. Vertically, on the left side of the screen, each word was written in Arabic orthography with no diacritics in the same order as they were played in the audio file (from the top: subject, verb, object, adverb). On the right side, there were squares confirming the ratings listeners gave for each word.

5.1.4 Data Analysis

The prominence ratings served as the dependent variable in mixed-effects regression models implemented in the program R (Baayen et al., 2008; R Core Team, 2014), using the lmer() function of the lme4 package (Bates, Maechler & Bolker, 2013). Models with different independent (or predictor) variables and random effects were constructed and compared with likelihood ratio tests using the ANOVA function. Variables that significantly improved model fit were retained in the best-fitting model. For these models, the independent variables always included item location (subject, verb, object), condition (matching, mismatching), focus location (subject, verb, predicate, object). In this chapter, the variable focus location is used to refer to the expected focus location based on the context. Models with maximal random effects structure failed to converge; therefore, a forward best-path method was used to determine which random slopes to include (Barr, Levy, Scheepers & Tily, 2013). As a result, only item location and condition (with no interaction between them) met the inclusion criterion (alph = .2), and thus, were included as random slopes with random intercepts for participants and items. The results from the best model are shown for each measure with parameter specific *p*-values obtained by using the Satterthwaite approximation, implemented in the lmerTest package in R (Kuznetsova et al., 2017).

5.2 Results

The results of the prominence ratings in the context-signal matching condition showed that focused words based on the context were perceived to be more prominent than unfocused words in the same utterance. Additionally, there is a declination trend in

the perceived prominence as we move from subject to object. That is, focused subjects were perceived to be more prominent than focused verbs, which in turn were perceived as more prominent than focused objects. Predicate focus showed a distributed prominence between both the verb and object, which were both perceived to be less prominent than the verb and object when they were individually focused.

For the context-signal mismatching condition, listeners saw different context questions, but the answers they heard after these questions were always the same answer with predicate focus. As discussed earlier, since this part of the experiment examines the independent effect of context on the perception of prosodic prominence, it was important that all acoustic information was held constant across the different contexts.

The prominence ratings for the context-signal mismatching condition closely resembled those from the context-signal matching condition, except when the set-up question asked about the subject. That is, when the context question asked about the subject in the context-signal mismatch condition, the subject was not perceived as the prominent word in the utterance. In the context-signal mismatching condition, the prominence ratings of utterances in the contextually focused subject show that all the words in the utterance had similar perceived prominence rating. However, the object was perceived to be more prominent. This pattern is compatible with object focus interpretation. The fact that the results from the context-signal mismatching condition generally resembled those from

the context-signal matching condition clearly demonstrates the effect that the context (set-up question in this case) had on the perception of prosodic prominence.

Table 17 below summarizes the average prominence rating divided by condition (Context-signal matching, Context-signal mismatching), item location (subject, verb, object), and focus location based on the question (subject, verb, object, predicate). In the context-signal mismatch condition, the entire predicate was always acoustically prominent. The predicate condition is always matching and is divided here merely for illustration purposes. These numbers are visualized in Figure 13 below.

Table 17
Average prominence ratings from perception experiment II.

Context focus location	Item location	Context-signal match		Context-signal mismatch	
		m	sd	m	sd
Subject	Subject	4.43	0.976	3.26	1.28
	Verb	2.88	0.992	3.11	1.22
	Object	2.64	0.939	3.46	1.42
Verb	Subject	2.92	1.21	2.83	1.21
	Verb	4.18	1.15	4.08	1.05
	Object	2.78	1.04	2.83	1.14
Predicate	Subject	2.83	1.24	2.91	1.31
	Verb	3.56	1.2	3.48	1.04
	Object	3.61	1.19	3.56	1.35
Object	Subject	2.99	1.26	2.91	1.27
	Verb	3.19	1.1	3.13	1.03
	Object	4.07	1.19	3.98	1.12

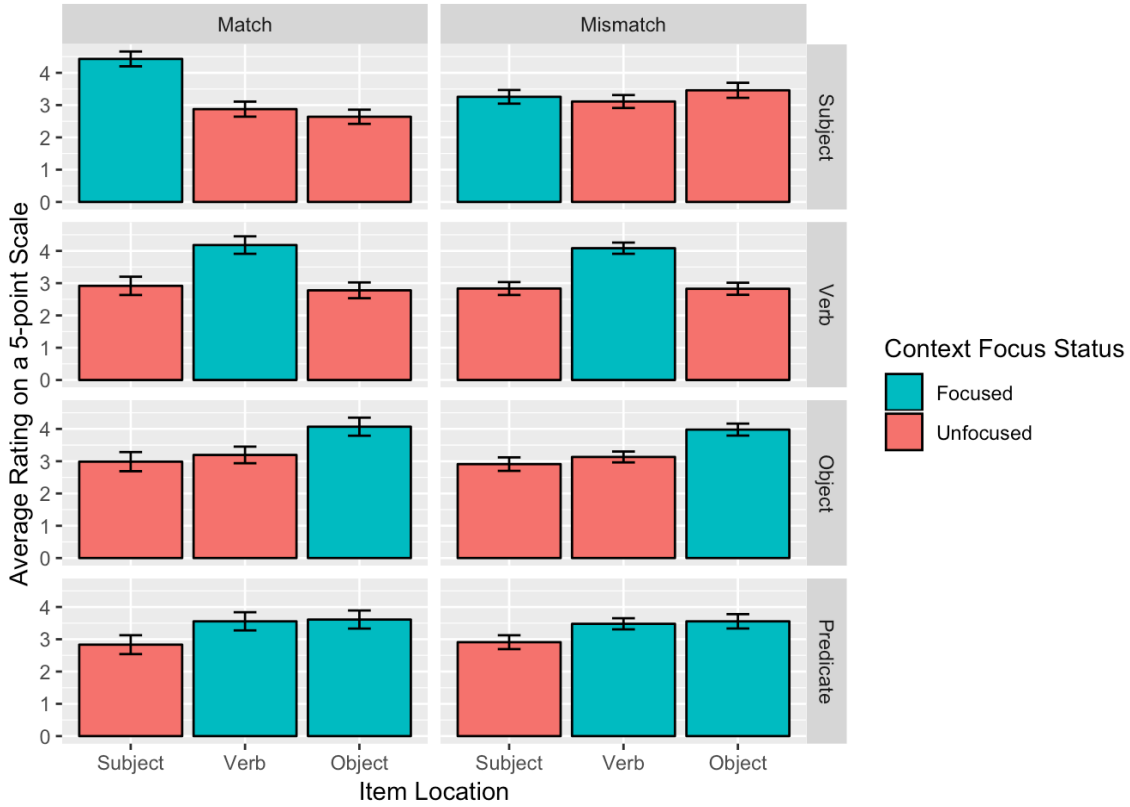

Figure 13 Perception experiment II: average prominence ratings A (error bars represent 95% confidence interval).

To test these patterns, a series of mixed-effects regression models were constructed in R (R Core Team, 2013) using the lmer() function of the lme4 package (Bates, Maechler & Bolker, 2013). The independent variables were condition (match, mismatch) focus location based on context (subject, verb, object, predicate) and item location (subject, verb, object). Condition and item location were included as random slopes with random intercepts for participants and items. The results from the best model are shown in Table 18 below with χ^2 statistics, standard error, *t*-value, and p-value.

Table 18
Coefficients of the best linear mixed-effects model of prominence ratings- experiment II (N= 2592).

	Model Summary			
	β	SE	*t*	*Pr(>|t|)*
Fixed effects				
(Intercept)	3.31	1.23	26.85	<0.**001**
Condition	-4.46	6.05	-0.73	0.49
Item location (V)	5.67	8.05	0.7	0.48
Item location (S)	-3.41	1.1	-3.09	**<0.01**
Focus location (V)	-1.03	4.88	-2.12	**<0.05**
Focus location (P)	6.94	4.23	0.16	0.86
Focus location (S)	-2.78	3.98	-0.69	0.48
Condition: Item location (V)	-1.81	8.96	-2.01	**<0.05**
Condition: Item location (S)	-4.05	7.76	-5.21	**<0.001**
Condition: Focus location (V)	3.24	1.04	0.31	0.75
Condition: Focus location (P)	4.17	8.96	0.46	0.64
Condition: Focus location (S)	6.94	8.45	0.08	0.93
Item location (V): Focus location (V)	2.16	1.2	18.09	**<0.001**
Item location (S): Focus location (V)	4.4	1.04	0.42	0.67
Item location (V): Focus location (P)	-2.94	1.04	-2.84	**<0.01**
Item location (S): Focus location (P)	-3.13	8.96	-0.34	0.72
Item location (V): Focus location (S)	-2.79	9.76	-2.86	**<0.01**
Item location (S): Focus location (S)	1.17	8.45	13.89	**<0.001**
Condition: Item location (V): Focus location (V)	-1.74	2.54	-0.68	0.49
Condition: Item location (S): Focus location (V)	-5.9	2.2	-0.26	0.78
Condition: Item location (V): Focus location (P)	3.82	2.2	0.17	0.86
Condition: Item location (S): Focus location (P)	1.72	1.9	0.9	0.36
Condition: Item location (V): Focus location (S)	-5.37	2.07	-2.59	**<0.01**
Condition: Item location (S): Focus location (S)	-1.73	1.79	-9.64	**<0.001**
	s^2			
Random effects				
Participant (Intercept)	0.52			
Item Location (Slope)	0.55			
Condition (Slope)	0.01			
Item (Intercept)	0.01			

To test the significance of these factors and their interactions, model comparison analyses were conducted. In the model comparisons, the full model (the model with the

best fit described above) was compared against the model without the effect in question. Accordingly, models with different predictors were computed and compared with likelihood ratio tests carried out by the ANOVA function, which used log-likelihood as a measure of goodness of fit (cf. Baayen, 2008). The model comparison analysis revealed a significant main effect of condition ($\chi^2(1) = 130.91$, $p < 0.001$, reflecting that there is a difference between the prominence ratings of sentences where the context and the signal were matching in terms of prosodic appropriateness and when the context and the signal were mismatching. A significant main effect of focus location based on the context was also found ($\chi^2(1) = 575.22$, $p < 0.001$), reflecting differences in the prominence ratings between the focus locations. However, this is specifically coming from the difference between the verb and object focus locations. A significant main effect of item location was also found ($\chi^2(1) = 606.55$, $p < 0.001$), reflecting that prominence ratings differed significantly across the three item locations. This difference obtains because the prominence ratings for items in the subject position were significantly lower than items in both the verb and object positions.

However, for the research question of this chapter, which examines whether naïve UNA listeners' expectations on how different aspects of the information structure are prosodically marked in production influence their perception of prosodic prominence, the interactions are more informative, particularly the three-way interactions among condition, item location and focus location. In terms of two-way interactions, model comparisons revealed a

significant interaction between condition and item location ($\chi^2(1) = 130.27$, $p < 0.001$). This difference comes from the subject-context of the mismatching condition, where words in the object position received significantly higher prominence ratings than verbs. Additionally, in the subject-context of the matching condition, subjects received higher prominence ratings than the verbs and objects. A significant interaction was also found between condition and focus location ($\chi^2(1) = 571.23$, $p < 0.001$). More importantly, however, the analyses revealed a significant three-way interaction between condition, focus location and item location ($\chi^2(1) = 596.01$, $p < 0.001$). This finding reflects that words in the verb position received significantly lower prominence ratings than the words in the object position when in the mismatching condition for the subject focus location. This is different from the pattern observed in the matching condition with the same parameter. Additionally, items in the subject position received higher ratings than the verb and object in the match condition under subject focus location than in the mismatch position. As mentioned earlier, the three-way interactions show that listeners responded similarly to both matching and mismatching conditions, except when the focus, as determined by the context question, was on the subject.

To summarize, listeners generally assigned higher prominence ratings to words that were expected to be focused based on the context than those that are expected to be unfocused. Relevant to the research question of this chapter is the three-way interactions between the condition, focus location based on the context question, and the item location. Accordingly, the three-way interactions tell us that listeners perceived and rated the prominence of words based on the context the same way between match and

mismatch conditions, except when the focus location according to the context was on the subject.

Taken together, the findings of this section indicate that, in the context-signal matching condition, as expected, listeners were able to successfully retrieve the information structures encoded by the speakers. Their performance was similar to the listeners' performance in the signal-based experiment. Additionally, in the context-signal mismatching condition, in which the context requires the focus to be on a different word than what is in the actual signal, an interesting pattern has emerged. When the context question asked about either the verb or object, listeners seem to perceive the verb or object to be prominent. However, when the context question asked about the subject, it did not elicit prominence perception on the subject.

Listeners seem to perceive prominence based on what the context is asking only when there is some kind of prominence in the signal to begin with. In the verb and object conditions, the actual signal had both the verb and object prominent since it was produced in response to predicate focus. However, the subject did not have any actual prominence in the signal itself.

When the context question asks about the subject in the mismatching condition, unlike what would be predicted by a context-based approach, listeners did not perceive the subject to be prominent, presumably because there was not enough acoustic evidence to support it.

5.3 Discussion

In this perception experiment, listeners were asked to listen and rate prominence on several utterances representing different information structures. The primary goal was to explore whether naïve listeners' reliance on contextual cues influences the perception of the information encoded by a speaker in an acoustic signal. That is, the question being asked is whether naïve UNA listeners' knowledge (or expectations) of how different aspects of the information structure are prosodically marked will influence their perception of prosodic prominence. If listeners' expectations on how information structures should be marked in a given context influence their perception of prosodic prominence, the results from the context-signal mismatching condition should be identical to the matching condition. This is because if context or expectations dictate perception, it will override the information in the speech signal. Additionally, if listeners' expectations or the context questions do not influence their perception of prosodic prominence, it is expected that the matching and mismatching conditions will be completely different. A third possibility, however, would be that listeners use a combination of contextual and acoustic cues to perceive prosodic prominence.

The trials in the matching condition represented a set-up question along with a prosodically appropriate answer to that question. When the question asked about the subject, the answer had only the subject focused in actual signal. When the set-up question asked about the verb, object or predicate, the answer had the appropriate prosodic realization to that question, by having the verb, object, and predicate focused, respectively. In this sense, the answers were genuinely produced in response to the

corresponding set-up question. In other words, the answers in the matching condition clearly marked focus by having the focused words prominent and unfocused words not prominent. So, in the subject focus, the subject was acoustically prominent but neither the verb nor the object was in that utterance. Similarly, when the verb was acoustically prominent neither the subject nor the object was prominent. This is also true for the object. The predicate focus had both the verb and the object acoustically prominent but not the subject.

The results of the matching condition mirrored the results from the signal-based perception experiment, which reflected the information encoded by the speakers in the first place. Listeners were highly successful in distinguishing the different focus locations in the matching condition, suggesting that UNA listeners were able to exploit acoustic cues to identify prominence. However, this does not truly tell us if context influenced the perception of prominence. Accordingly, the mismatching condition was included to explore that question and the matching condition allowed for a direct comparison between the two.

In the mismatching condition, the set-up questions asked about words in the subject, verb, object positions but the answer following these questions was the one originally produced in response to a predicate focus question, so that all the acoustic information was held constant across all the trials. The predicate was chosen as the uniform answer because the verb and object are both acoustically prominent and pronounced in one intonational phrase. This means listeners listened to the same answer after different set-up questions. This allowed us to see

if context influenced their perception. As mentioned earlier, if listeners are perceiving prominence based on the context, then no difference should be expected between matching and mismatching conditions. However, if they are responding to the signal alone then their prominence ratings should be the same across the different focus locations within the mismatching condition because they are responding to the same acoustic signal.

The results from the mismatching condition demonstrated two different patterns. First, when the set-up question (or context) asked about the verb or the object, listeners' ratings reflected perceived prominence consistent with a context-based perception (i.e., not consistent with what is in the acoustic signal). That is, when the set-up question asked about the verb (so the focus is expected to be on the verb in the answer), but the answer has both the verb and the object focused. Since a predicate focus utterance is being used, participants heard the verb to be prominent but not the object (even though the object was also acoustically prominent in the answer).

Similarly, when the set-up question asked about the object, but the acoustic signal in the answer had both the verb and object focused, listeners heard prominence on the object but not on the verb. This finding shows evidence that context (or expectation based on the context) affected listeners' perception of prominence. Further, the context did not only affect the perception of the target word so that it is perceived as prominent, but also words that were not required by the context to be prominent received low prominence scores. Increasing the prominence score on the target word and decreasing the prominence score on the surrounding words maximizes perceived prominence. This

pattern is consistent with an expectation-based perception. If listeners are primed by the context to anticipate a specific prosodic organization, then the rating would be expected to reflect the most prototypical behavior expected in production. That is, in the production, speakers tend to maximize the distinction between prominent and non-prominent words by maximizing the acoustic correlates on the target item and minimizing them on the surrounding items. This is what seems to happen in this case.

However, when the subject was expected to be focused based on the set-up question (context), but the answer did not have any prominence on the subject, but rather on the predicate, listeners did not hear the subject to be prominent compared to the other words in the utterance. Participants significantly rated the object as the most prominent word in that utterance, which is the pattern one expects in object focused utterance. This finding shows that listeners were generally not responding to their expectations. These observations are represented in Figure 14 below.

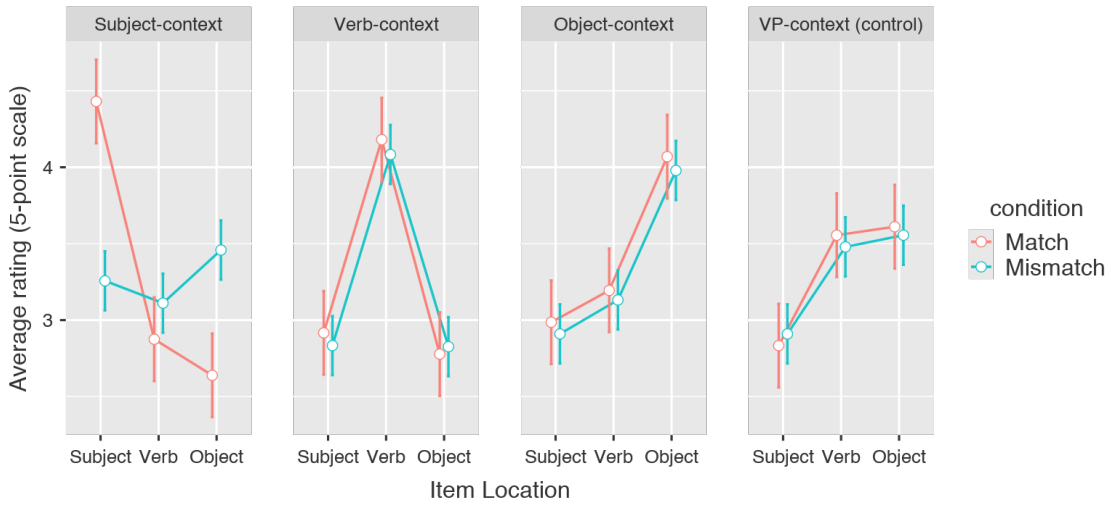

Figure 14 Perception experiment II: average prominence ratings B. (Panels represent the fours question context (subject, verb, object, verb phrase), Colors represent condition (match, mistmatch), error bars represent 95% confidence interval).

In addition to the previous observations, Figure 14 shows that in the subject-context panel, the subject received slightly higher prominence rating in the mismatch condition than subjects in the other contexts, even though the subject was never acoustically focused in any of these mismatched conditions. This suggests that in the mismatch condition, under the subject-context, listeners perhaps started off initially based on their expectations (as seems to be required by the context). The fact that words in subject position were perceived to have somewhat prominent support a context-based interpretation. However, the fact that listeners rated the object to be more prominent reflects some information found in the speech signal. In fact, in the subject context, in the mismatch condition, the subject was never acoustically prominent, and the rating seems

to reflect that it is not acoustically prominent. The verb also was perceived not to be prominent in this condition.

Additionally, by comparing the verb and object prominence ratings in the mismatch condition in the verb and object contexts to the verb and object in the VP context, it is apparent that the context further enhanced the strength of prominence perception. To put this more clearly, the control stimuli in the mismatch condition were the predicate trials. That is, in the mismatch condition, the VP-context and its answers were actually "matching." The same predicate focus answer was used after the other focus contexts (i.e., subject, verb, object). Accordingly, this allows for a direct comparison between the different contexts within the same mismatch condition. In this case, the verbs were perceived to have higher prominence in the verb-context condition than it was perceived in the control VP-context condition.

Similarly, the object was perceived to have higher prominence ratings in the object-context condition than the VP-context condition. In addition, listeners did not only hear the verb to be prominent in the verb-context condition, but they also heard the object to be less prominent. Listeners also did not only hear the object to be the most prominent word in the utterance but also heard the verb to be less prominent. These ratings were of the same audio file. This pattern could only be found if listeners were responding to their expectations, and not to the signal.

These patterns in the mismatching condition beg the question of why listeners seemed to rely on expectation-based cues in the verb and object contexts

but seem to rely more on signal-based cues in the subject context even though the auditory stimuli are identical across the different contexts.

A possible explanation could be related to the accent distribution in the answer. As previously mentioned, in the mismatch condition, the same audio file was used for all the contexts. The accent distribution on the actual audio file had the subject un-accented, and the verb and object accented; this is because the audio file was originally produced in response to a predicate focus. Accordingly, in the verb and object contexts, listeners did not hear any prominence on the subject because neither the context nor the acoustic signal has it focused. The verb and object are prominent in the acoustic signal, so when the context is asking for one to be prominent, listeners heard it as such, and simply ignored the other accent. Further, the context enhanced the prominence on the verb or object in their respective contexts by not only causing the focused words to be perceived as more prominent but also by causing the contextually unfocused words to be perceived as less prominent. However, in the subject context, the conflict between the context and the actual speech signal was more pronounced and obvious because there was no actual accent on the subject, to begin with, suggesting that for contextual cues to be operational, acoustic cues need to facilitate such process.

It could be that listeners perceive more prominence on a word that it is acoustically prominent if the context motivates it but ignore prominence on it if the context is not asking for it. However, when there is no accent on a word, to begin with, listeners do not have the trigger to consider it, especially when other words in the utterance are acoustically prominent. In other words, listeners seem to be able to ignore

an accent, but they do not presume one even when motivated by the context cue. This is shown by the fact that listeners were able to ignore the accents on the verb or object when the context did not ask about them but could not on the subject. The findings from this experiment show that listeners use a combination of both signal-based and context-based cues. Although the results show that context influences the perceived prominence, it seems to be acoustically conditioned. If contextual factors act independently from acoustical factors, then the perception of prominence would be context-based, and the context would completely override information found in the signal. However, this is not what is truly observed. It may seem that context cues override signal cues in most cases, but these cases seem to require some acoustic prominence for context cues to operate. This last point is demonstrated by the disparity between the perceived prominence in the subject versus object and verb conditions. It seems that the perception of prosodic prominence is both signal-based and context-based. Further, it seems that signal-based cues are an essential component for contextual cues to modulate the perception of prominence.

Chapter 6

General Discussion and Conclusion

The three experiments reported in this dissertation have explored how different aspects of information structure are conveyed with prosody and how listeners use acoustic and contextual cues to perceive encoded information. In the first experiment, naïve speakers engaged in a production task in which they heard context questions and read the answer sentences in SVO word order. The set-up questions were used to manipulate the focus status in the target answers. The target sentences were acoustically analyzed in terms of F0, intensity, and duration. The results from the production study demonstrate that UNA speakers used F0, intensity, and duration to mark different information structures.

In the second experiment, naïve listeners listened to the target utterances from the previous production experiment without any context and rated the prominence of each word, using a 5-point prominence rating scale. The goal of the first perception experiment was to examine whether naïve listeners could perceive the prosodic prominence encoded by the speakers from the speech signal alone. The results demonstrated that listeners were successful in perceiving acoustic information encoded by the speakers' speech signal alone, as reflected by their prominence rating.

In the third experiment, naïve listeners saw a context question and heard an answer to that context question. The question and the answer were sometimes matching (i.e., the answer was prosodically appropriate to the question) and sometimes mismatching (i.e., the answer was prosodically inappropriate to the question). The listeners' task was to read the question and rate the prosodic prominence on every word in the answer. The goal of this experiment was to explore whether listeners' expectations on how the information structures should be marked (as specified by the context) affect their perception of prosodic prominence. The results provided evidence that listeners' expectations affected their perception of prosodic prominence but also showed that listeners resorted to the acoustic signal in extreme cases where the conflict between the expectation and the actual signal could not be overlooked. (In this case, it was the lack of pitch accent).

This dissertation asked four questions about the prosodic marking of prominence (as a consequence of focus) and the perception of such prominence. In the following section, these four questions are discussed and answered in turn.

6.1 Research Questions Revisited

6.1.1 RQ1: Do Speakers of UNA Mark Information Structure Prosodically in SVO Utterances, and to What Extent?

The results from the production study have shown that UNA speakers prosodically mark focus as part of the information structure in a very systematic way. First, UNA speakers explicitly mark focus location across the different sentential positions (subject, verb, object) using acoustic cues. They also mark the size (or breadth)

of the focus systematically. UNA speakers prosodically differentiate focus on a smaller element (e.g., object focus) from wider focus (e.g., predicate focus). They do so by producing more prominence on the object in the object only focus rather than when the object was focused as part of a predicate focus, suggesting that perhaps speakers are aware of the possible ambiguity that could arise if the object was marked with similar prominence levels in these two conditions. The results also show that UNA speakers do not overwhelmingly differentiate between contrastive from noncontrastive (narrow) focus.

Though the distinction between contrastive and noncontrastive narrow focus has been argued to exist theoretically, previous literature has not always been successful in showing that such distinction has an acoustic consequence. A couple of factors may have contributed to this weak prosodic differentiation between these contrastive and noncontrastive narrow categories.

First, speakers in this study were not told of the possible ambiguity of their production between contrastive and noncontrastive narrow answers and were possibly not aware of such ambiguity. Results from Breen et al., (2010) show that English speakers differentiated between contrastive and noncontrastive focus only after they were aware of the possible prosodic ambiguity. Second, in this experiment, UNA speakers read the answers from a computer screen. The nature of this task may have contributed to this weak prosodic differentiation. Moreover, while the results showed that speakers only used F0 cues to differentiate contrastive from noncontrastive focus and that the patterns of these two categories were the same, they do not provide enough evidence to suggest

that they are marked differently in the language. Therefore, whether UNA speakers are able to differentiate between contrastive and noncontrastive sentences remains an open question.

6.1.2 RQ2: What are the Acoustic Features Associated with the Prosodic Prominence of the Different Information Structures in SVO Utterances in UNA?

The acoustic and statistical analyses of the production data revealed that UNA speakers use duration, maximum intensity, maximum F0, and F0 range to mark focus across different locations and different focus domains. The data show that the stressed vowel of the focused word is associated with longer duration, higher maximum F0, wider F0 range, and higher intensity than unfocused words. These acoustic correlates of prosodic prominence in UNA are consistent with previous findings on other spoken Arabic varieties.

6.1.3 RQ3: In the Absence of a Discourse Context, How Well Can Naïve UNA Listeners Retrieve Different Aspects of the Information Structure Intended by the Speaker from the Speech Signal?

In the first perception experiment, listeners were able to rate prosodic prominence of words in utterances in a way reflecting the actual structure of information in which the sentence was originally produced. For example, when the sentence [θa:mir ha:waʃ sa:rah ʔilju:m] *Thamer scolded Sarah today* was produced in response to a question like *who scolded Sarah today?* the acoustic analysis revealed that the subject was the only prominent word in the sentence, the verb, and object were not prominent. Accordingly,

when listeners rated this sentence in isolation, the ratings reflected that they perceived the subject to be the most prominent and the other words not prominent.

With regard to focus location, listeners were always successful in rating the subject, verb, object, and predicate to have the highest prominence in the sentence, when the target sentence was originally produced with the focus on the subject, verb, object, or predicate, respectively. In terms of the breadth of the focus, listeners were also able to distinguish among broad focus, narrow focus on the subject, verb or object, and predicate focus. However, listeners were not able to reflect a difference between contrastive and noncontrastive productions, perhaps because the distinction between the two categories was small and negligible.

6.1.4 RQ4: Does Naïve UNA Listeners' Knowledge of How Different Aspects of the Information Structure are Marked Prosodically Influence Their Perception of Prosodic Prominence?

In the second perception experiment, listeners show evidence that their judgment was affected by their expectations or by the context. The similarities in the prominence ratings between the matching and mismatching conditions are quite revealing. This experiment also shows that listeners may have started with an expectation-based approach but then resorted to trusting in the speech signal in some cases. Listeners were not only able to ignore and suppress existing prosodic prominence on words because the context motivated it, but they also imagined the prominence on the target word to be even more prominent, and this was motivated by the context. However, when the context caused the prominence to be on a word that was not acoustically prominent, listeners

were faced with a serious conflict and eventually seem to have resorted to information from the speech signal to resolve the conflict.

6.2 Theoretical Implications

One of the issues this dissertation has examined is whether contrastive and noncontrastive focus are differentiated prosodically by UNA. This issue is related to two open questions in the literature regarding *contrast*, one phonological and one semantic. First is the question of whether there is a phonological category [contrast]. Second, is whether the terms contrastive and noncontrastive constitute separate semantic/pragmatic categories.

The findings of this dissertation do not provide enough support for a separate phonological category [contrast]. If speakers differentiate the two, we would expect phonetic differences in their realization. In the production experiment, UNA speakers did not notably differentiate between utterances in contrastive and noncontrastive conditions. Moreover, the contrastive and noncontrastive focus had similar patterns. Thus, explicit contrast (in the contrastive condition) did not result in a substantial difference between the two focus conditions, namely contrastive and noncontrastive utterances. The differences were only found in the F0 results in the production, and these differences were minimal. Additionally, listeners' ratings of contrastive and noncontrastive productions did not differ between the two categories. This is supported by the fact that no statistically significant difference was found between the ratings of contrastive and noncontrastive utterances in the perception experiment. Perhaps since the difference in

terms of F0 values between contrastive and noncontrastive found in the production, is not enough to be noticeable by the listeners.

These findings are also informative for the debate over whether contrastive and noncontrastive focus constitute separate semantic categories. According to Rooth (1992), contrastive and noncontrastive do not constitute separate semantic/pragmatic categories. The findings of this dissertation perhaps offer support to Rooth's account, especially in light of the phonetic evidence that these two do not seem to be categorically differentiated. However, future studies need to explore this issue further.

6.3 Methodological Considerations

This dissertation included production and perception experiments to evaluate the questions at hand. In the production experiment, the acoustic analysis was used to evaluate multiple phonetic cues objectively. Further, in the perception experiments, different approaches to speech perception were tested (i.e., signal-based analysis (bottom-up) and context-based analysis (top-down.) This allowed for a direct comparison between the two approaches and helped us in answering the research questions. Additionally, the perception experiments efficiently employed an online survey to collect real-time judgment using a modified RPT task. The methodological considerations described in the methods sections of chapters 3,4 and 5 could be beneficial when designing similar studies.

6.4 Typological Contributions

Given that the production and perception of prosodic prominence in UNA is experimentally understudied, this dissertation adds to the existing literature on Arabic by

providing findings for a wider typological comparison among Arabic varieties, as well as cross-linguistically. Findings from the production experiment demonstrated that UNA speakers use F0, intensity, and duration to mark focus prosodically.

6.5 Conclusion and Future Directions

This dissertation has explored important questions related to the production and perception of focus as marked by prosodic prominence. The findings provide valuable insights about UNA, as well as the relationship between prosody and information structure in production and perception. Nonetheless, many questions need to be explored. Future research needs to specifically explore the production and perception of contrastive sentences. For example, how do UNA speakers mark prominence in a more communicative task? When UNA speakers are made aware of the possible ambiguity between contrastive and noncontrastive, how is this knowledge acoustically reflected, and consequently, perceived by the listeners? The results from the context-based experiment have revealed that prosodic prominence may be affected by acoustic and non-acoustic factors, but it is not clear yet how these factors are weighted and consolidated during processing.

Appendix A. List of Experimental Sentences

Sentence 1			
	Condition	Elicitation Question	Target answer
1	broad focus	wiʃ sˁar ʔilju:m what happened today?	
2	narrow focus [s]	mi:n ha:waʃ sa:rah ʔilju:m who scolded Sarah today?	
3	narrow focus [v]	θa:mir wiʃ sawa ʔil sa:rah ʔilju:m what did Thamer do to Sarah today?	
4	narrow focus [o]	θa:mir ha:waʃ mi:n ʔilju:m Whom did Thamer scold today?	(laʔ) θa:mir ha:waʃ sa:rah ʔilju:m
5	narrow focus [vp]	θa:mir wiʃ sawa ʔilju:m What did Thamer do today?	(no) Thamer scolded Sarah today
6	contrastive focus [s]	mi:n ha:waʃ sa:rah ʔilju:m xali:d who scolded Sarah today? Khaled?	
7	contrastive focus [o]	θa:mir ha:waʃ mi:n ʔilju:m danah Whom did Thamer scold today? Danah?	
8	contrastive focus [v]	θa:mir wiʃ sawa ʔil sa:rah ʔilju:m samaħha What did Thamer do to Sarah today? forgave her?	
9	contrastive focus [vp]	θa:mir wiʃ sawa ʔilju:m wasˁal ɣadah What did Thamer do today? dropped off Ghadah?	

Sentence 2			
	Condition	Elicitation Question	Target answer
1	broad focus	wiʃ sˁar ʔilba:riħ what happened last night?	
2	narrow focus [s]	mi:n sa:maħ ri:ma ʔilba:riħ who forgave Rima last night?	
3	narrow focus [v]	sa:lim wiʃ sawa ʔil ri:ma ʔilba:riħ what did Salim do to Rima last night?	
4	narrow focus [o]	sa:lim sa:maħ mi:n ʔilba:riħ whom did Salim forgive last night?	(laʔ) sa:lim sa:maħ ri:ma ʔilba:riħ
5	narrow focus [vp]	sa:lim wiʃ sawa ʔilba:riħ what did Salim do last night?	(no) Salim forgave Rima last night
6	contrastive focus [s]	mi:n sa:maħ ri:ma ʔilba:riħ ħamad who forgave Rima last night? Hamad?	
7	contrastive focus [o]	sa:lim sa:maħ mi:n ʔilba:riħ rawa:n whom did Salim forgive last night? Rawan?	
8	contrastive focus [v]	sa:lim wiʃ sawa ʔil ri:ma ʔilba:riħ ha:waʃha what did Salim do to Rima last night? scolded her?	
9	contrastive focus [vp]	sa:lim wiʃ sawa ʔilba:riħ kalam ħisˁah What did Salem do last night? talked to Hessah?	

Sentence 3			
	Condition	Elicitation Question	Target answer
1	broad focus	wiʃ sˁar ʔilju:m what happened today?	
2	narrow focus [s]	mi:n xa:sˁam di:na: ʔilju:m who fought with Dina today?	
3	narrow focus [v]	ʕa:mir wiʃ sawa ʔil di:na: ʔilju:m what did Amer do to Dina today?	
4	narrow focus [o]	ʕa:mir xa:sˁam mi:n ʔilju:m whom did Amer fight today?	(laʔ) ʕa:mir xa:sˁam di:na: ʔilju:m
5	narrow focus [vp]	ʕa:mir wiʃ sawa ʔilju:m what did Amer do today?	(no) Amer fought with Dina today
6	contrastive focus [s]	mi:n xa:sˁam di:na: ʔilju:m xa:lid who fought Dina today? Khalid?	
7	contrastive focus [o]	ʕa:mir xa:sˁam mi:n ʔilju:m rawa:n whom did Amir fight today? Rawan?	
8	contrastive focus [v]	ʕa:mir wiʃ sawa ʔil di:na: ʔilju:m wa:sˁalha what did Amer do to Dina today? dropped her off?	
9	contrastive focus [vp]	ʕa:mir wiʃ sawa ʔilju:m qa:bl nu:r What did Amer do today? met Noor?	

Sentence 4			
	Condition	Elicitation Question	Target answer
1	broad focus	wiʃ sˁar ʔams what happened yesterday?	
2	narrow focus [s]	mi:n ra:sal li:na ʔams who texted Lina yesterday?	
3	narrow focus [v]	sa:mi: wiʃ sawa ʔil li:na ʔams what did Sami do to Lina yesterday?	
4	narrow focus [o]	sa:mi: ra:sal mi:n ʔams whom did Sami text yesterday?	(laʔ) sa:mi: ra:sal li:na ʔams
5	narrow focus [vp]	sa:mi: wiʃ sawa ʔams what did Sami do yesterday?	(no) Sami ra:sal Lina yesterday
6	contrastive focus [s]	mi:n ra:sal li:na ʔams fahad who texted Lina yesterday? Fahad?	
7	contrastive focus [o]	sa:mi: ra:sal mi:n ʔams ħan:n whom did Sami text yesterday? Hanan?	
8	contrastive focus [v]	sa:mi: wiʃ sawa ʔil li:na ʔams kallamha: what did Sami do to Lina yesterday? Talked to her?	
9	contrastive focus [vp]	sa:mi: wiʃ sawa ʔams ra:ħ ʔil su:q What did Sami do yesterday? Went to the mall?	

Sentence 5			
	Condition	Elicitation Question	Target answer
1	broad focus	wiʃ sˁar ʔilba:riħ what happened last night?	
2	narrow focus [s]	mi:n ma:zaħ ja:ra ʔilba:riħ who joked (with) Yara last night?	
3	narrow focus [v]	na:jif wiʃ sawa ʔil ja:ra ʔilba:riħ what did Naif do to Yara last night?	
4	narrow focus [o]	na:jif ma:zaħ mi:n ʔilba:riħ whom did Naif joke (with) last night?	(laʔ) na:jif ma:zaħ ja:ra ʔilba:riħ
5	narrow focus [vp]	na:jif wiʃ sawa ʔilba:riħ what did Naif do last night?	(No) Naif joked (with) Yara last night
6	contrastive focus [s]	mi:n ma:zaħ ja:ra ʔilba:riħ_fahad who joked with Yara last night Fahad?	
7	contrastive focus [o]	na:jif ma:zaħ mi:n ʔilba:riħ ħan:n whom did Naif joke with last night? Hanan?	
8	contrastive focus [v]	na:jif wiʃ sawa ʔil ja:ra ʔilba:riħ kallamha: what did Naif do to Yara last night? Talked to her?	
9	contrastive focus [vp]	na:jif wiʃ sawa ʔilba:riħ ra:ħ ʔil su:q What did Naif do last night? Went to the mall?	

Sentence 6			
	Condition	Elicitation Question	Target answer
1	broad focus	wiʃ sˁar ʔilju:m what happened today?	
2	narrow focus [s]	mi:n dʒa:mal nu:rah ʔilju:m who complemented Norah today?	
3	narrow focus [v]	ra:mi: wiʃ sawa ʔil nu:rah ʔilju:m what did Rami do to Norah today?	
4	narrow focus [o]	ra:mi: dʒa:mal mi:n ʔilju:m whom did Rami complement today?	(laʔ) ra:mi: dʒa:mal nu:rah ʔilju:m
5	narrow focus [vp]	ra:mi: wiʃ sawa ʔilju:m what did Rami do today?	(No) Rami complemented Norah today
6	contrastive focus [s]	mi:n dʒa:mal nu:rah ʔilju:m sa:mir who complemented Norah today? Samer?	
7	contrastive focus [o]	ra:mi: dʒa:mal mi:n ʔilju:m ya:ra whom did Rami complement today? Yara?	
8	contrastive focus [v]	ra:mi: wiʃ sawa ʔil nu:rah ʔilju:m ha:waʃha: what did Rami do to Norah today? scolded her?	
9	contrastive focus [vp]	ra:mi: wiʃ sawa ʔilju:m ʃa:f sa:lih What did Rami do today? saw Saleh?	

Appendix B. Language Background Questionnaire

1. How old are you? (must be over 18 years of age)
2. Gender:
3. Where were you born?
    a. City:
    b. Country:
4. What is your native language?
5. Education:
6. Do you speak any other languages?
7. When did you start learning your second language and how?
8. Have you lived in a country where English is a native language?
9. How long have you lived in a country where English is a native language?
10. Do you have any known hearing or speaking problems?

References

Almalki, H. & Morrill, T. (2016). Yes/No question intonation in Urban Najdi Arabic. *Speech Prosody 2016*: 606-610.

Almalki, H. (2016). Prosodic focus in Urban Najdi Arabic. *Unpublished qualifying paper*, George Mason University.

Alshamari, M. (2017). *Topic particles in the North Hail dialect of Najdi Arabic*. Ph.D Thesis, Newcastle University.

Alzaidi, M. (2014). *Information Structure and Intonation in Hijazi Arabic.* Ph.D Thesis. Department of Language and Linguistics, University of Essex.

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

Bakir, M. J. (1979). *Aspects of clause structure in Arabic; a study in word order variation in literary Arabic*. Bloomington, IN: Indiana University.

Baltazani, M., & Jun S- A. (1999). Focus and topic intonation in Greek. *Proceedings of the XIVth International Congress of Phonetic Sciences*, vol. 2 (San Francisco, USA): 1305 – 1308

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of memory and language*, *68*(3), 255-278.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). *lme4: Linear mixed-effects models using Eigen and S4*. R package version 1.1-10.

Birner, B. J. (1994). Information status and word order: An analysis of English inversion. *Language*, 233-259.

Bishop, J. (2012). Information structural expectations in the perception of prosodic prominence. In G. Elordieta & P. Prieto (Eds.), *Prosody and Meaning*. Berlin, Boston: DE GRUYTER. https://doi.org/10.1515/9783110261790.239

Bock, J. K., & Mazzella, J. R. (1983). Intonational marking of given and new information: Some consequences for comprehension. *Memory & Cognition*, *11*(1), 64–76. https://doi.org/10.3758/BF03197663

Bolinger, D. L. (1961). Contrastive Accent and Contrastive Stress. *Language*, *37*(1), 83. https://doi.org/10.2307/411252

Breen, M., Fedorenko, E., Wagner, M., & Gibson, E. (2010). Acoustic correlates of information structure. *Language and Cognitive Processes*, *25*(7–9), 1044–1098. https://doi.org/10.1080/01690965.2010.504378

Buring, D. (2010). Towards a typology of focus realization. Malte Zimmermann and Caroline Féry (eds.), *Information Structure*, 177–205

de Jong, K., & Zawaydeh, B. A. (1999). Stress, duration, and intonation in Arabic word-level prosody. *Journal of Phonetics*, *27*(1), 3–22. https://doi.org/10.1006/jpho.1998.0088

Halliday, M. A. K. (1967). Notes on transitivity and theme in English Part I. *Journal of Linguistics*, *3*(01), 37. https://doi.org/10.1017/S0022226700012949

Boersma, P., & Weenink, D. (2015) Praat: doing phonetics by computer, (Version 5.4.22) http://www.praat.org/

Chafe, W. (1976). Givenness, contrastiveness, definiteness, subjects, topics and point of view. In Charles N. L. (ed.), *Subject and Topic*. (27-55). New York: Academic Press

Chahal, D. (2001). *Modeling the intonation of Lebanese Arabic using the autosegmental-metrical framework: a comparison with English*. Unpublished PhD thesis, University of Melbourne

Chahal, D., & Hellmuth, S. (2014). The intonation of Lebanese and Egyptian Arabic. *In Prosodic Typology II.* Jun, S- A. (Ed). Oxford: Oxford University Press

Clark, H. H., & Haviland, S. E. (1977). Comprehension and the given-new contract. *Discourse production and comprehension. Discourse processes: Advances in research and theory*, *1*, 1-40.

Cole, J., Mo, Y., & Hasegawa-Johnson, M. (2010). Signal-based and expectation-based factors in the perception of prosodic prominence. *Laboratory Phonology*, *1*(2). https://doi.org/10.1515/labphon.2010.022

Cole, J., & Shattuck-Hufnagel, S. (2016). New Methods for Prosodic Transcription: Capturing Variability as a Source of Information. *Laboratory Phonology*, *7*(1), 8. https://doi.org/10.5334/labphon.29

Eady, S. J., & Cooper, W. E. (1986). Speech intonation and focus location in matched statements and questions. *The Journal of the Acoustical Society of America*, *80*(2), 402-415.

Fery, C. (2007). The prosody of topicalization. In Schwabe, K., & Winkler, S. (eds.), *On Information Structure, Meaning and Form: Generalizations across Languages*. Amsterdam: John Benjamins, 69-86.

Gussenhoven, C. (1983). Focus, mode and the nucleus. *Journal of Linguistics*, 19, 377-417.

Gut, U., & Bayerl, P. S. (2004). Measuring the reliability of manual annotations of speech corpora. *Proceedings of Speech Prosody 2004, 565-568*

Hartmann, K., & Zimmermann, M. (2009). Morphological focus marking in Gùrùntùm (West Chadic). *Lingua*, *119*(9), 1340–1365. https://doi.org/10.1016/j.lingua.2009.02.002

Haruna, A. (2003). *A Grammatical Outline of GùrdùN/Gùrùntùm (Southern Bauchi, Nigeria)* (Vol. 25). Köppe.

Hellmuth, S. (2006). *Intonational pitch accent distribution in Egyptian Arabic*. Unpublished PhD thesis, SOAS.

Hirotani, M., Frazier, L., & Rayner, K. (2006). Punctuation and intonation effects on clause and sentence wrap-up: Evidence from eye movements. *Journal of Memory and Language*, *54*(3), 425-443.

Holes, H. (2010). Word order and textual function in Gulf Arabic. In: Owens, J. & Elgibali, A., (eds.), *Information Structure in Spoken Arabic*. London; New York: Routledge, pp. 61-74.

Ingham, B. (2010). Information structure in the Najdi dialects. In: Owens, J. & Elgibali, A., (eds.), *Information Structure in Spoken Arabic*. London; New York: Routledge, pp. 75-92.

Ito, K., Speer, S., & Beckman, M. (2004). Informational status and pitch accent distribution in spontaneous dialogues in English. *Proceedings of Speech Prosody 2004*, 279-282.

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

Jun, S.-A. (2012). Prosodic Typology Revisited: Adding Macro-rhythm. *Proceedings of Speech Prosody, 6th international conference*, Shanghai, China.

Jun, S.-A. (2014). *Prosodic typology: the phonology of intonation and phrasing*, Oxford University Press

Kiss, K. (1998). Identificational focus versus information focus. *Language*, 74: 245–273

Krifka, M. (2008). Basic notions of information structure. *Acta Linguistica Hungarica*, *55*(3–4), 243–276. https://doi.org/10.1556/ALing.55.2008.3-4.2

Kihm, A. (1999). Focus in Wolof. *The grammar of focus*, *24*, 245.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: tests in linear mixed effects models. Journal of Statistical Software, 82(13).

Ladd, D. (2008). *Intonational Phonology*. Cambridge: Cambridge University Press.

Lambrecht, K. (1994). *Information structure and sentence form: Topics, focus, and the representations of discourse referents*. Cambridge: Cambridge University Press.

McCarthy, J. J. (1979). On stress and syllabification. *Linguistic inquiry*, *10*(3), 443-465.

Moutaouakil, A. (1989). *Pragmatic Functions in a Functional Grammar of Arabic*. Dordrecht: Foris.

Musabhien, M. (2009). Case, agreement and movement in Arabic: a minimalist approach. PhD Thesis: Newcastle University.

Neeleman, A., Titov, E., van de Koot, H., & Vermeulen, R. (2009). A syntactic typology of topic, focus and contrast. In J. Van Craenenbroeck (ed.) *Alternatives to Cartography*. Berlin: Mouton de Gruyter, 15 – 51

Peirce, J. W. (2007). PsychoPy - Psychophysics software in Python. *Journal of Neuroscience Methods,* **162** (1-2):8-13 doi:10.1016/j.jneumeth.2006.11.017

Pierrehumbert, J., & Hirschberg, J. (1990). The meaning of intonational contours in the interpretation of discourse. In *Intentions in communication*, (eds.) Cohen, P., James Morgan, J., & Pollack,M., 271 - 311. Cambridge MA: MIT Press.

Pitt, M.A., Dilley, L., Johnson, K., Kiesling, S., Raymond, W., Hume, E. and Fosler-Lussier, E. (2007) Buckeye Corpus of Conversational Speech (2nd release) [www.buckeyecorpus.osu.edu] Columbus, OH: Department of Psychology, Ohio State University

Qualtrics, I. (2018). Qualtrics. *Provo, UT, USA*.

R Core Team, (2014). *R: A language and environment for statistical computing*, Vienna: R Foundation for Statistical Computing.

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

Rump, H., & Collier, R. (1996). Focus conditions and the prominence of pitch - accented syllables. *Language and Speech*, 39, 1-17

Sityaev, D., House, J. (2003). Phonetic and phonological correlates of broad, narrow and contrastive focus in English. *Proc. of the 15th ICPhS*, Barcelona, 1819-1822.

Stalnaker, R., (1974). Pragmatic presuppositions. In: Munitz, M.K., Unger, Peter, K. (Eds.), *Semantics and Philosophy*. New York University Press, New York.

Stalnaker, R. (2002). Common ground. *Linguistics and philosophy*, *25*(5-6), 701-721.

Soltan, U. (2007). *On formal feature licensing in minimalism: Aspects of Standard Arabic morphosyntax* (Doctoral dissertation).

Szabolcsi, A. (1981). Compositionality in focus. *Folia Linguistica*, 15(1-2), 141-162.

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

Turnbull, R., Royer, A. J., Ito, K., & Speer, S. R. (2017). Prominence perception is dependent on phonology, semantics, and awareness of discourse. *Language, Cognition and Neuroscience*, *32*(8), 1017–1033. https://doi.org/10.1080/23273798.2017.1279341

Watson, J. C. E. (2011). Word stress in Arabic. In M. Oostendorp, C. Ewen, E. Hume, & K. Rice (Eds.), *The Blackwell companion to phonology*. Volume 5 (pp. 2990 - 3018). Oxford: Blackwell.

Xu, Y., Chen, S.-w., & Wang, B. (2012). Prosodic focus with and without post-focus compression (PFC): A typological divide within the same language family? *The Linguistic Review*, 29: 131-147

Xu, Y., & Xu, C. (2005). Phonetic realization of focus in English declarative intonation. *Journal of Phonetics*, 33: 159-197

Biography

Hussain Almalki received his Bachelor of Arts in English language and literature from Al-Imam Muhammad ibn Saud University, Riyadh, Saudi Arabia, in 2010. He earned his Master of Arts in Linguistics, in 2014, from Florida International University, Miami, FL.