PRODUCTION AND PERCEPTION OF INTONATIONAL CUES EXPRESSING EPISTEMIC CONFIDENCE IN TURKISH

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF INFORMATICS OF
THE MIDDLE EAST TECHNICAL UNIVERSITY
BY

AYŞENUR HÜLAGÜ

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE
IN
THE DEPARTMENT OF COGNITIVE SCIENCE

AUGUST 2018

Approval of the thesis:

PRODUCTION AND PERCEPTION OF INTONATIONAL CUES EXPRESSING EPISTEMIC CONFIDENCE IN TURKISH

Submitted by AYŞENUR HÜLAGÜ in partial fulfillment of the requirements for the degree of **Master of Science in Cognitive Science Department, Middle East Technical University** by,

Prof. Dr. Deniz Zeyrek Bozşahin
Dean, **Graduate School of Informatics** _____

Prof. Dr. Cem Bozşahin
Head of Department, **Cognitive Science** _____

Assist. Prof. Dr. Umut Özge
Supervisor, **Cognitive Science Dept., METU** _____

Assoc. Prof. Dr. Annette Hohenberger
Co-Supervisor, **Cognitive Science Dept., METU &**
Institute of Cognitive Science, University of Osnabrück _____

Examining Committee Members:

Prof. Dr. Deniz Zeyrek Bozşahin
Cognitive Science Dept., METU _____

Assist. Prof. Dr. Umut Özge
Cognitive Science Dept., METU _____

Assoc. Prof. Dr. Annette Hohenberger
Cognitive Science Dept., METU &
Institute of Cognitive Science, University of Osnabrück _____

Prof. Dr. Salih Selçuk İşsever
Linguistics Dept., Ankara University _____

Assist. Prof. Dr. Duygu Özge
Foreign Language Education Dept., METU _____

Date: _____

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last name : Ayşenur Hülagü

Signature : _____

ABSTRACT

PRODUCTION AND PERCEPTION OF INTONATIONAL CUES EXPRESSING EPISTEMIC CONFIDENCE IN TURKISH

Hülagü, Ayşenur

MSc., Department of Cognitive Science

Supervisor: Assist. Prof. Dr. Umut Özge

Co-Supervisor: Assoc. Prof. Dr. Annette Hohenberger

August 2018, 140 pages

This study investigates how certainty and uncertainty are conveyed and comprehended among Turkish interlocutors. A production study was conducted to elicit utterances under epistemic modalities of "certain" and "uncertain" (in which the participants uttered only locative pronouns ("burada" (here), "şurada" (there), and "orada" (there)), only epistemic adverbs ("galiba" (probably) and "kesinlikle" (certainly)), and locative pronouns modified by epistemic adverbs). Speakers tended to extend the duration to express uncertainty and to shorten the duration to express certainty. In terms of pitch range, when speakers were uttering locative pronouns and expressing that they were sure of their answers, they used a wider pitch range. A perception study was conducted to observe the effects of acoustic properties (i.e. mean pitch, pitch range, and duration), intonation contours and lexical items on the perception of epistemic confidence of the speaker. The task was choosing modality of utterance on the two alternative forced-choice method and a 7-point Likert scale in a within-subjects design. Among the acoustic properties the effect of duration was found to be prominent. Shorter durations refered to higher certainty whereas longer durations refered to higher uncertainty. Moreover, high pitch range determined confidence of speakers as certain. In terms of intonation contours, perception of the listeners suggested that H*+L LL% is the best representer of all categories. Duration helped to distinguish epistemic confidence. The other representative intonation contours are: H* LL% for certainty and L* LH% for uncertainty.

Keywords: Intonation Contour, Speaker Confidence, Modality, Locative Pronoun, Epistemic Adverb

ÖZ

TÜRKÇEDE BİLGİ GÜVENİ İFADE EDEN EZGİSEL İPUÇLARININ ÜRETİMİ VE ALGISI

Hülagü, Ayşenur

Yüksek Lisans, Bilişsel Bilimler Bölümü

Tez Yöneticisi: Dr. Öğr. Üyesi Umut Özge

Eş danışman: Doç. Dr. Annette Hohenberger

Ağustos 2018, 140 sayfa

Bu çalışmada, kesinlik ve kesinlik dışılığın Türkçe iletişimde nasıl aktarıldığı ve anlaşıldığı araştırılmaktadır. "Kesin" ve "kesinlik dışı" bilgi kipleri ile (sadece işaret zamirleri ("burada", "şurada" ve "orada"), bilgi zarfları ("galiba" ve "kesinlikle") ile bilgi zarfları ile nitelenen işaret zamirleri katılımcılar tarafından söylenerek bir üretim çalışması gerçekleştirilmiştir. Katılımcılar kesinlik dışılığı ifade ederken süreyi artırma, kesinliği ifade ederken ise süreyi azaltma eğilimi gösterdiler. Perde aralığı açısından konuşurlar işaret zamirlerini söylerken cevaplarından emin olduklarını ifade ettikleri zaman geniş perde aralığı kullandılar. Akustik özelliklerin (ortalama ses perdesi, perde aralığı ve süre), ezgi konturu ve leksik öğelerin konuşurun bilgi güveninin algılanışı üzerindeki etkilerini gözlemlemek için bir algı çalışması yapıldı. Görev, iki alternatifli mecburi seçim yönteminde ve denek içi dizayn ile 7 dereceli Likert ölçeğinde söylemin kipliğini seçmektir. Akustik özellikler arasında sürenin etkisi belirgin olarak bulundu. Daha kısa süreler daha yüksek kesinliğe işaret ederken, daha uzun süreler daha yüksek kesinlik dışılığa işaret etti. Geniş perde aralığı konuşurun güvenini kesin olarak belirledi. Ezgi konturu açısından dinleyicilerin algısı, H*+L LL% 'nin tüm kategorilerde en iyi temsilci olduğunu ortaya koydu. Süre, bilgi güvenini ayırt etmede yardımcı oldu. Diğer temsilci ezgi konturları kesinlik için H* LL% ve kesinlik dışılık için L* LH% olarak belirlendi.

Anahtar Sözcükler: Ezgi Konturu, Konuşurun Güvenilirliği, Kiplik, İşaret Zamiri, Bilgi Zarfı

Off to a new chapter,

ACKNOWLEDGMENTS

First and foremost, I would like to thank my advisors Assoc. Prof. Dr. Annette Hohenberger and Assist. Prof. Dr. Umut Özge for their extraordinary support in thesis process. Their willingness to give their time has been very much appreciated. I am indebted to them for their understanding, patience, wisdom, enthusiastic encouragement, and for showing me I could do more than I thought.

The completion of this thesis could not have been possible without assistance of my committee. I would like to express my sincere gratitude to Prof. Dr. Deniz Zeyrek Bozşahin for not letting me give up and believing in me. I would like to thank Prof. Dr. Salih Selçuk İşsever and Assist. Prof. Dr. Duygu Özge for their valuable comments. Their contributions are sincerely appreciated and gratefully acknowledged.

I would like to thank especially to Ayşegül Özkan for sharing her statistical expertise with me and teaching patiently.

Special thanks to my beloved friends; Buket İkizoğlu, Özgen Demirkaplan, and Tuğçe Katrancı Plaza for always being next to me even if they are far far away from me.

Another special thanks to my lovely team; Burcu Alapala, Seyfullah Demir, and Sinan Onur Altınuç for their company along the way and sharing all the stress from this year.

To all of my friends who supported me, either emotionally and physically, thank you.

I also would like to thank the speakers of speech elicitation experiment and participants of speech perception experiment for their wonderful collaboration.

Last but not least, I would like to thank my loving grandmother Meryem Müberrra Boduroğlu and my loving aunt Serra Mübeccel Gültürk with a special feeling of gratitude for their endless support.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AM	Autosegmental-Metrical Model
ANOVA	Analysis of Variance
C.Dur	Centered Duration
CMPitch	Centered Mean Pitch
CPRange	Centered Pitch Range
LMM	Linear Mixed Model
M	Mean
METU	The Middle East Technical University
Msec	Millisecond
R	The R Project for Statistical Computing
SD	Standard Deviation
Sec	Second
ToBI	Tones and Break Indices

CHAPTER 1

INTRODUCTION

Mankind is in need of distributing their knowledge via language. The role of language in our social life is not only limited to describing facts but it is also utilized to make an impact on the others that we interact with (Ergenç, 2002). Smith & Clark (1993) described the function of language in social interaction as transfer of information that also includes the belief state of the speaker. Agents such as speaker and listener, and context bring about speech acts that carry meaning (Green, 1979). The components of communication are source, message, and target. In our study, these components correspond to the following roles. The speaker is the source of the message. Confidence of the speaker (also "speaker certainty") is to what extent the speaker considers the content of his/her utterance as true. The act of uttering aims to convey both the content of the message and speaker confidence concerning this content; therefore an addressee is targeted and required to grasp both the meaning and the speaker confidence. Comprehension of speaker confidence leads to more efficient communication.

1.1 Background

The category of epistemic modality has the values certain and uncertain. Certainty is the belief of a speaker encoded within the utterance. In our study we used the term "certainty" in a condition in which speaker has the relevant information. On the other hand, the term "uncertainty" refers to the lack of or weakness of information. In such a condition when the speaker is certain, their confidence is named "certain". This is because in this condition the speaker is sure of what is uttered. In the other condition when the speaker is uncertain, their confidence is named "uncertain". This is because in this condition speaker is not sure of what is uttered. The linguistic devices used to convey speaker confidence, namely epistemic modality, have different contributions to pragmatics.

There is also a relationship between intonation and pragmatics (Barth-Weingarten, 2009). Intonation serves as a device to indicate information structure, mental state of the speaker, and turn taking. There are various studies that put importance on the relation of intonation to speaker confidence. The perception of speaker confidence was examined through a task in which speakers answered questions about facts. Then, these answers are rated by the participants as stimuli. The participants were more successful to detect cues of only one type of cue (auditory or visual) (Swerts & Krahmer, 2005).

The relationship between speaker confidence and lexical items has also been the focus of research. Moore, Bryant, and Furrow (1989) studied the development of these terms in children which signal certainty and also reveal the mental state of the children. The lexical devices that contribute to the relationship between pragmatics and speaker confidence can be divided into two: epistemic adverbs and locative pronouns. Lexical words which are employed in increasing the force of a speech act are "certainly", "definitely", "wihout doubt", etc. whereas the ones employed in decreasing the force of a speech act are "perhaps", "probably", "doubtful", etc. (Holmes, 1982).

Moreover, there have also been some studies that investigated the interaction of lexical and intonational cues to the expression of speaker confidence. Borràs-Comes, Roseano, Vanrell, Chen, & Prieto (2011) showed that the effect of lexical items might be surpassed by prosody and gestures. Speech perception is more likely to be affected by auditory cues rather than visual cues (Dijkstra, Krahmer, & Swerts, 2006). A wide range of studies showed that adults place importance on auditory cues rather than lexical cues in terms of affective communication (see Argyle, Alkema, & Gilmour, 1971; Morton & Trehub, 2001; Reilly & Muzekari, 1986; Solomon & Ali, 1972, as cited in Friend, 2003). Furthermore, Gravano, Benus, Hirschberg, German, & Ward (2008) concluded that speaker confidence requires the modal verb "would" and pitch track showing a downstepped contour.

A final factor that might be relevant for the expression and perception of speaker confidence is "cognitive accessibility" which is defined as "the ease (of effort) with which particular mental contents come to mind" (Kahneman, 2003, p. 699). Piwek, Beun, & Cremers (2007) show that cognitive accessibility is closely related to physical-proximity, whose linguistic indicator is demonstrative pronouns. Construal level theory, on the other hand, claims that people construe events as well as objects differently in terms of their spatio-temporal distance from themselves (Trope & Liberman, 2010). According to construal level theory, psychological distance refers to epistemic distance to mental presentation of an event or object.

1.2 The Present Study

The present study was inspired by the work of Hübscher, Esteve-Gibert, Igualada, & Prieto (2017) and Jiang & Pell (2017). We wanted to answer questions such as "How do speakers convey epistemic confidence?" and "What would be the role of speech in the evaluation of epistemic confidence?" Under the light of previous findings regarding the topic, we aim to investigate the phonological, lexical and spatial dimensions of indication and perception of speaker confidence. For this purpose we used Turkish as a case domain. For the investigation of phonological aspects we used the tonal analyses of recorded elicited utterances (mean pitch, pitch range, intonational contours); for the investigation of lexical aspects, we used different epistemic adverbials (kesinlikle 'certainly', galiba 'probably'); and for the investigation of spatial factors we used demonstrative pronouns forming a scale of

proximity (burada ‘here’, şurada ‘there’, orada ‘there’). We were also interested in the interaction of these potential components.

We conducted a production and perception study. In the production study, we created an imaginary scenario where some objects are hidden to various locations and the speaker is asked to answer questions about the location of the hidden object with varying degrees of speaker confidence. Our aim was to observe how speakers regulate the acoustic properties (i.e., mean pitch, pitch range, and duration) during speech and whether this regulation will be affected by lexical items such as locative pronouns (burada, şurada, orada) and epistemic adverbs (kesinlikle, galiba) under certain and uncertain conditions.

The perception study was conducted in order to examine the comprehension of the listeners, whether the certainty of the speaker is determined by the above mentioned properties and items. The question was the contributions of acoustic properties for the perception of epistemic confidence. We want to determine the relation between prosodic cues and pragmatic meaning. Specifically, we seek to determine how differences among prosodic elements such as mean pitch, pitch range, and duration and intonational contour of an utterance in interaction with epistemic adverbs and locatives affect perception of speaker confidence (i.e., certain and uncertain).

We expect from our study to provide new findings about the expression of cognitive states. Epistemic confidence of the speaker will be conveyed with distinctive acoustic features through intonation, locative pronouns, and adverbs in speech addressed to adults. We investigate the communicative aspect of conveyance of speaker confidence. Smith & Clark (1993) pointed out that uncertainty cues such as intonation, tempo, and pausing are communicable and also received by listeners in order to be used to differentiate between confidence categories of the speaker. If listeners are able to predict epistemic confidence of the speaker using only acoustic features, this finding would indicate that uncertainty and certainty are not accessible via locative pronouns. Moreover, if listeners are able to predict the epistemic confidence of the speaker using the lexical information of epistemic adverbs, this finding would indicate that adults take notice of lexical meaning to resolve uncertainty and certainty difference rather than making use of acoustic features.

The results of this study will shed light on how speakers manipulate both acoustic properties and lexical items to convey their message and to which extent it is possible to observe the effect of these manipulations on the comprehension of the addressee. The advantage of having the knowledge of these effects for the listeners can result in engagement with more accurate understanding that leads to more effective communication. To be able to determine the effect of non-verbal and verbal cues in encoding and decoding of epistemic confidence would help us to suggest a cognitive model for acoustic correlates of mental imagery which is the imagined scenario where the key is. It is important to notice the effect of inferences from utterances of speakers regarding their confidence during social interaction.

Therefore, it is necessary to determine which lexical cues and acoustic properties are predictive for speaker confidence.

In essence, our experiments can lead the way for future research in the field of prosody and more specifically on the topic of epistemic confidence of speakers. This topic is related to self-expression, cognition, and, in a broader sense, human interaction. This study would also provide support for studies on the development of comprehension in infants by investigating distinctive feature of intonation that convey the mental state of the speaker (Burenhult, 2003; Hübscher, Esteve-Gibert, Igualada, & Prieto, 2017; Jiang & Pell, 2017; Küntay & Özyürek, 2006; Piwek, Beun, & Cremers, 2007).

A literature review on prosodic features and their pragmatics, intonation contours, pragmatics of speaker confidence, of intonation contours, of lexical items and studies on speaker confidence will be provided in Chapter 2. Chapter 3 and Chapter 4 will describe the methods used in both production and perception experiments. Also, several analyses will be explained that were run to predict the factors contributed to the participant responses. Chapter 5 will mention commonalities as well as discrepancies observed in intonation by providing examples from the data. Chapter 6 will discuss the findings from the data as well as attempt to present evidence to support them. Finally, Chapter 7 will summarize the results, touch on limitations, and suggest future work.

CHAPTER 2

LITERATURE REVIEW

This chapter aims to introduce the basic concepts which are addressed and referred to throughout the thesis.

2.1 Prosodic Features and Pragmatic Meaning

Hirst & Di Cristo (1998) have illustrated a schema in which prosodic characteristics are classified. In the first place phonetic is divided into two main types: “Cognitive” (phonological) and “Physical” (acoustic). To begin with cognitive type, “lexical prosodic systems” and “non-lexical prosodic systems” are included in this type. Stress, tone, and quantity are the elements of “lexical prosodic systems”. Regarding “non-lexical prosodic systems”, intonation proper is the only element. Next, physical type is comprised of “prosodic parameters”. Fundamental frequency, intensity, duration, and spectral characteristics are these parameters. These systems and parameters contribute to intonation (p. 7).

Intonation is defined as “the body of tone changes” during uttering (Ergenç, 2002, p. 64). For Ladd (1996), definition of intonation does not involve stress, accent, tone, “paralinguistic features” (i.e., tempo and loudness) yet they affect each other. Intonation is used for “suprasegmental phonetic features” that are composed of fundemental frequency (F_0), intensity, and duration to carry “postlexical” intentions (Ladd, p. 6). Balog & Brentari (2008) described suprasegmental aspects that constitute prosody in production of speech. These aspects are pitch, lenght, and loudness. Acoustic correlates of these features are fundamental frequency, duration, and intensity, respectively. With respect to “postlexical”, intonation acts on utterances or speech acts by carrying intentions (pp. 7-8). On the other hand, the definition by Hirschberg & Pierrehumbert (1986) for intonational features includes phrasing, accent placement, pitch range, and tune. They used the term tune to refer to intonation contour which is composed of pich accents, phrase accents, and boundary tones. These features will be used to characterize the intonation of utterances in the thesis.

Intonation can be used to emphasize on the prominence of an utterance with respect to information structure (Nolan, 2014). The works of Pierrehumbert & Hirschberg (1990), Steedman (1991) and Vallduvı´ (1992) provided how to approach information structure and pointed out components of an utterance which is

incumbent upon release of information (as cited in Özge & Bozşahin, 2010). The components are disassemblied with respect to mental representations of interlocutors, referents, and knowledge states (Arnold, Kaiser, Kahn, & Kim, 2013).

2.1.1 Fundamental Frequency (F_0) and Pitch

Fundamental Frequency (F_0) is a physical property with the measure of Hertz (Hz) and defined as "the number of times the vocal folds open and close in one second" (Simpson, 2009, p.623). The terms pitch and F_0 are used to refer to the same. While F_0 is defined as a "physical" feature, pitch is depended on F_0, and defined as its "pyschophysical" aspect (Ladd, 1996, p. 7). As the frequency increases the voice becomes acute with high-pitch and as the frequency decreases the voice becomes grave with low-pitch. Due to stress on the words the frequency increases and resulted in high-pitched voice (Demircan, 2001). The studies found out that high pitch is employed in anger and surprise (Uldalli, 1964; Williams & Stevens, 1972, as cited in Ladd, 1996, p. 12).

Pragmatic aspect of different pitch values are mentioned by Balog & Brentari (2008). Psychoacoustic saliency is ascribed to rising contour due to frequency-following-response (FFR) proposed by Krishnan, Xu, Gandour, & Cariani (2004) (as cited in Chen, Zhu, & Wayland, 2017). When epistemic confidence of the speaker about an utterance is certain that is expressed with falling pitch at the end (Childs, 2014). While replying questions how much the time spent on finding answer depends on belief state of the speaker. If belief state is more likely to be close to certain, search time would diminish (Smith & Clark, 1993). With respect to changes in pitch such as rising or falling, shorter duration of the stimuli predicted categorical perception of listeners of tonal languages. Also, a tonal language such as Mandarin has a tendency of categoric perception rather than non-tonal languages such as English. The reason is sensitivity towards pitch contour changes (Chen, Zhu, & Wayland, 2017).

Changes in pitch contour can change the semantic of the same utterance in a non-tonal language. A falling pitch contour is signal for a statement whereas a rising pitch contour is signal for a question (Chen, Zhu, & Wayland, 2017). Falling pitch indicates that an utterance is completed by speaker and addressee's turn to come. On the other hand, stable or rising pitch value indicates that flow of information is still inact (Martha, 1996). In order to introduce new topics, utterances start with high intonation. On the other hand a start with low intonation indicates continuity. With respect to ending of the utterances continuity is related to high boundary tone whereas finality is related to low boundary tone (Gussenhoven, 2002).

On the other hand, another acoustic property that is pitch height is made use of in Catalan to distinguish question types whether wh- or yes-no questions due to H tone (Vanrell, Mascaró, Prieto, & Torres-Tamarit, 2009, as cited in Borràs-Comes, Roseano, Vanrell, Chen, & Prieto, 2011). Low ending resulted in an utterance to be perceived as statement while high ending resulted in an utterance to be perceived as question based on evaluation of Catalan participants (Borràs-Comes, del Mar

Vanrell, & Prieto, 2014). In an intonation study by Hadding-Koch & Studdert-Kennedy (1964), Swedish and English listeners were required to differentiate utterances and indicate whether they were statements or questions. The results suggested that perception of question correlates with higher peak and and pitch.

Emotions such as sadness, calmness, and security are associated with low pitch value whereas emotions such as, anxiety and joy are associated with high pitch values (Rodero, 2011). High and/or rising fundamental frequency conveys politeness and lack of confidence while low and/or falling fundamental frequency conveys aggression and confidence (Bolinger, 1978; Ohala, 1984).

Among acoustic features duration and pitch work together (Nolan, 2014). Lehiste (1976) proposed that dynamic changes in F_0 values cause misperception of duration (as cited in Yu, Lee, & Lee, 2014). Although durations are equal, utterances with wider F_0 values can be perceived as longer than utterances with narrower F_0 values (Yu, 2010). Rising pitch contour, high pitch, high vowel are perceived as utterance has longer duration (Yu, 2010; Yu, Lee, & Lee, 2014). The type of vowels such as high or low vowels influences perception of duration in a sense that a high vowel extends duration (Yu, Lee, & Lee, 2014).

2.1.2 Pitch Range

The difference of pitch value between the maximum and the minimum pitch value gives *pitch range*. Pitch range is affected by speakers' use of their voice (raising or falling). The aim of speakers for manipulating their pitch range (increase or decrease in pitch range) might be to emphasize on the utterance (Pierrehumbert & Hirschberg, 1990) (see Figure 2.1).

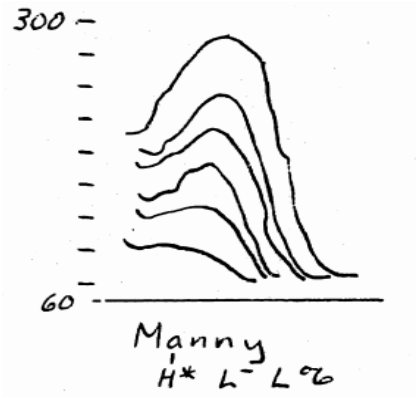

Figure 2.1: The utterance with H* LL% intonation contour in six different pitch ranges. Adapted from "The phonology and phonetics of English intonation," by J. B. Pierrehumbert (1980), Doctoral dissertation, Massachusettes of Technology, p. 269.

Within the literature of intonational phonology, pragmatic aspects of pitch range in different languages remains doubtful (Borràs-Comes, del Mar Vanrell, & Prieto, 2014). Free Gradient Hypothesis explains pragmatic function of widen pitch range as

to emphasize on the utterance (Ladd, 1996). Pitch range encourages the meaning of a statement and help establishing it clearly in speech (Borràs-Comes, Roseano, Vanrell, Chen, & Prieto, 2011). The dynamic change of pitch helped to distinguish between confidence and unconfidence. With respect to fundamental frequency range, the utterances which were evaluated as confident had highest value (Jiang & Pell, 2017). Employment of pitch range can be manipulated in order to create an effect on the addressees by speakers. This manipulation can be exemplified as the use of narrow pitch range to comfort and the use of wider pitch range to allert the addressee.(Fernal, 1989).

Pitch range serves different pragmatic functions in different languages for instance to distinguish question types. Within the language of Bari Italian in which pitch range is made use of to distinguish question types whether information-seeking or echo questions. Pitch range is narrower for information-seeking and statements than echo questions (Borràs-Comes, Roseano, Vanrell, Chen, & Prieto, 2011; Savino & Grice, 2007). In English, pitch range can cause a change in the categoric perception of an utterance for instance from uncertainty to incedulity (Hirschberg & Ward, 1992). As regarding pitch range studies by Face (2005, 2007, 2011) in Spanish showed that pitch value in the first peak in an utterance helped for categorical perception of utterances as declaratives and questions. Since pitch value is greater in the first peak of questions, wider pitch range can be associated with questions. Thus, declaratives have narrower pitch range (as cited in Borràs-Comes, del Mar Vanrell, & Prieto, 2014).

2.1.3 Stress

Stress is defined as force of breath on syllable in an utterance. In Turkish focus is marked by stress (Göksel & Özsoy, 2000). Also, stress and intonation are in close relationship in English as well. Pitch accent corresponds to stressed syllable (Nolan, 2014). Stress is observed in the last syllable of a lexical thus Turkish is considered stress-accent language (Lewis, 1970). Since Turkish is one of intonation languages, the location of the stressed syllable in the word conveys different meanings (Demircan, 2001). Stress in Turkish carries information related to speaker and perception of speaker (Ipek & Jun, 2014). Stress is signalized through duration (Pierrehumbert & Hirschberg, 1990). Loudness and high pitch are considered relevance of stress by Kornfilt (1997). Demirezen (2014) suggested that phoneme with pitch accent associated with stress and high intonation on this phoneme.

2.1.4 Accent

Accent is defined as stress on one syllable within the utterance. In Catalan, pitch accent can be distinctive for statements and echo questions. Borràs-Comes, Vanrell, & Prieto (2010) grouped pitch accents for statements and echo questions. L+H* is employed in statements whereas L+¡H* is employed in echo questions. "¡" indicates upstepped rising. In Turkish it is found at the end of the utterance (Ergenç, 2002). Speakers aim to make the addressee to believe in the same by manipulating their

intonation in order to convey the meaning. This is the reason for choosing different accent types.

2.1.5 Tone

Tone is defined as the level of frequency. The meaning can be distinguished by tone (Ergenç, 2002).

Gandour (1977) pointed out that vowels with high tones are more likely to have shorter durations than low tones. There are not many studies examined the relation between duration and high/low level tones (see Faytak & Yu, 2011).

2.2 Intonation Contours

The pattern of F_0 contour is determined by the tones L (low) and H (high) and these tones produce tune. If a stressed syllable has one of these tones, it is *pitch accent*. With respect to F_0, stress can be cued by both lower pitch range L* and higher pitch range H*. Moreover, pitch accents which are composed of stressed tones and tones can be exemplified as H*+L, H+L*, L*+H, and L+H*. Another unit of intonation contour is *boundary tone* which can be H or L tone and is located at the end of an utterance. Moreover, boundary tones which are composed of two tones and a diacritic '%' can be exemplified as HH%, HL%, LL%, LH% (Pierrehumbert, 1980; Pierrehumbert & Hirschberg, 1990).

Pierrehumbert (1980) in her thesis provided a grammar which is a "finite state" as the following. The grammar gives all the possible combinations of intonational elements (see Figure 2.2).

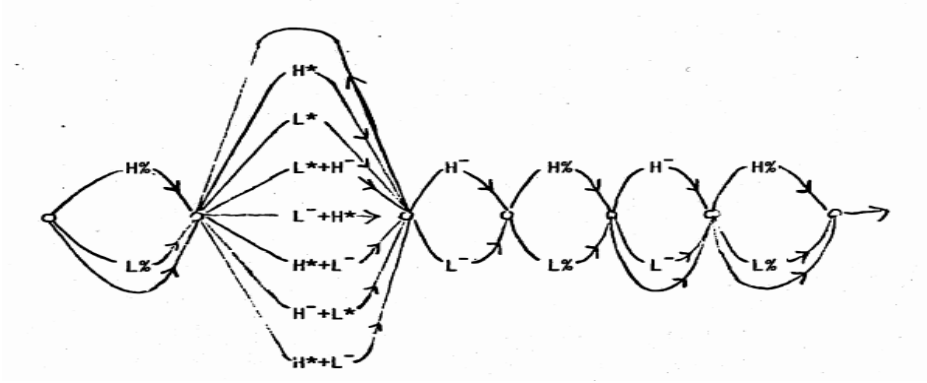

Figure 2.2: Grammar of an intonational contour (boundary tone, pitch accents, phrase accent, and boundary tone) in English intonation. Adapted from "The phonology and phonetics of English intonation," by J. B. Pierrehumbert (1980), Doctoral dissertation, Massachusettes of Technology, p. 29.

2.3 ToBI

ToBI stands for tones and break indices. Autosegmental-Metrical Model (AM) has a transcription system which is ToBI and allows for intonation studies. It is considered as a language system for spoken language (Pierrehumbert, 1980)[1]. Initially, it was set for American English (see Figure 2.3), today there are adaptations for different languages such as German, Dutch, and Catalan (Nolan, 2014).

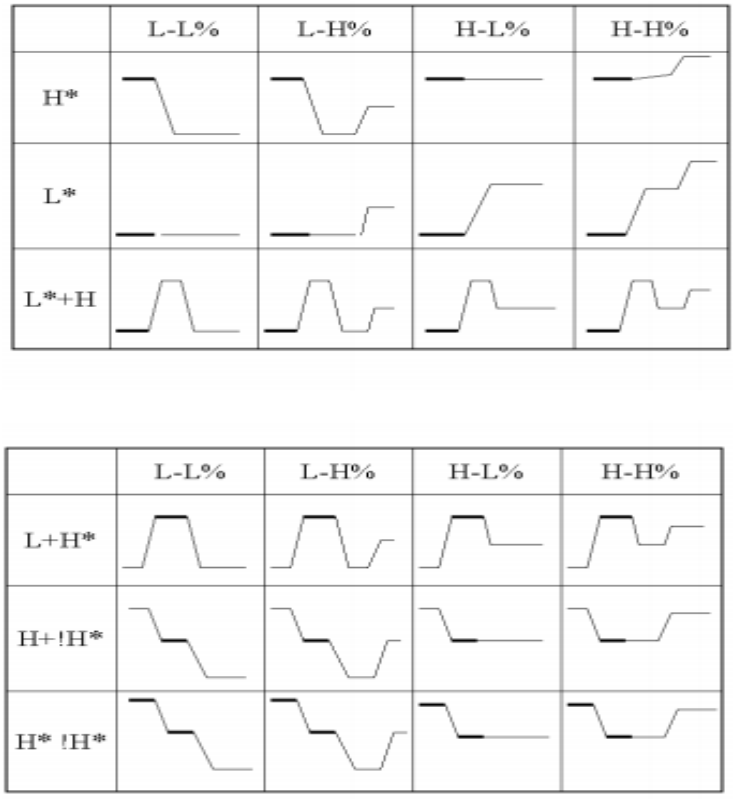

Figure 2.3: ToBI contours for Standard American English. Adapted from "Pragmatics and Intonation," by J. Hirschberg (2004), In *The Handbook of Pragmatics,* p. 10.

2.4 Types of Intonation Contours

Tracks are to depict the definitions of intonation which are "physical" and provided below (Ladd, 1996, p. 11). Pierrehumbert (1980) examined intonation contours in English language and presented pitch tracks of them. Some example intonation contour types that occured in our data as well depicted below.

---

[1] The exclamation mark (!) indicates "downstep" in which the tone is the same (L or H) but pitch level is higher for preceding syllable. However, "downstep" does not occur in our data.

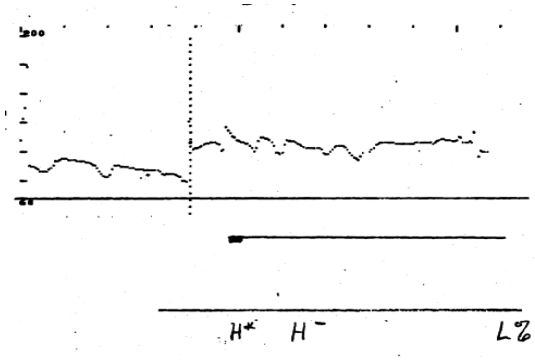

Figure 2.4: H* HL% intonation contour. Adapted from "The phonology and phonetics of English intonation," by J. B. Pierrehumbert (1980), Doctoral dissertation, Massachusettes of Technology, p. 392.

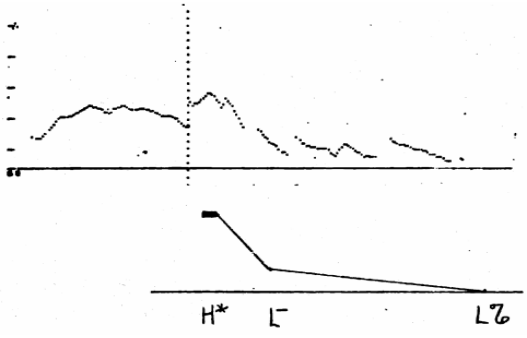

Figure 2.5: H* LL% intonation contour. Adapted from "The phonology and phonetics of English intonation," by J. B. Pierrehumbert (1980), Doctoral dissertation, Massachusettes of Technology, p. 391.

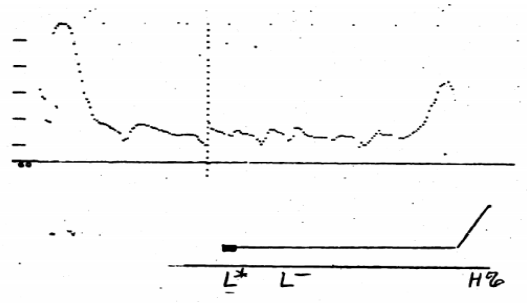

Figure 2.6: L* LH% intonation contour. Adapted from "The phonology and phonetics of English intonation," by J. B. Pierrehumbert (1980), Doctoral dissertation, Massachusettes of Technology, p. 394.

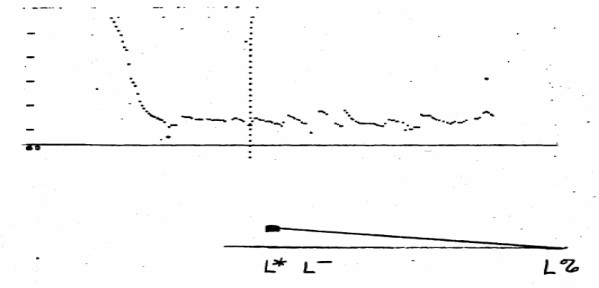

Figure 2.7: L* LL% intonation contour. Adapted from "The phonology and phonetics of English intonation," by J. B. Pierrehumbert (1980), Doctoral dissertation, Massachusettes of Technology, p. 394.

2.5 Pragmatic Meaning and Speaker Confidence

The term 'speech act' is generally understood to mean illocution and they are used interchangeably (Searle, 1976). The term is explained by acts such as asking, asserting, etc. In these speech act examples, uttering brings along speaker meaning. While the idea that is held by Grice (1968) emphasizes the purpose of affecting the addressee, it has been challenged by a suggestion that speaker meaning is a token of mental state of an individual (Green & Williams, 2007).

Leech and Svartvik (1975) have divided communication into classes in terms of meaning. Pragmatic meaning, in other words interactional meaning, builds one of these classes. This type of meaning places importance on interlocutors equally. Both speaker and addressee are featured with respect to their mental state. According to a classification by Holmes (1982), this class corresponds to affective meaning due to having speaker's aim of change in addressee by the utterance. Thus, speaker confidence is highlighted. Speakers have the degree of commitment which is relative to their speech. The term speaker confidence interprets to what extent the speaker is committed to what has been uttered (Palmer, 2001). For Leech and Svartvik (1975), logical meaning is another class of communication which is related to factual conditions. Logical meaning corresponds to modal meaning in Holmes' (1982) classification. Mental state of a speaker in respect of factual conditions of the utterance is featured within this type of meaning. This makes reference to epistemic modality. Epistemicity is related to belief of the speaker about the proposition. The definition of epistemic stance can be broadened to include degree of knowledge of speaker and confidence to what the one uttered. A Speaker can convey modality by three means of expression: lexical (i.e., certainly, probably, etc.), syntactic (i.e., may, must, etc.), and morphological (i.e., declarative, subjunctive, etc.). These three means of expression help to grab epistemic stance between interlocutors.

According to a definition provided by McNeill (1992), the term analogical signal is used to refer to prosody (voice pitch and loudness). Indeed, these signals are paralinguistic-beyond language. These analogical or paralinguistic signals are sometimes used to convey epistemic modality. Holmes (1982) has provided a

definition of epistemic modality with regard to speaker. The term is used to state the degree of confidence to the utterance token about its factual condition which is conveyed by means of expression. When certainty or uncertainty is conveyed, epistemic modality is implicated. Certainty, belief, and doubt regarding the utterance are conveyed by the speaker (Palmer, 2014). In this way, speaker confidence can be resolved by the analysis of the utterance and its factual condition.

2.6 Pragmatic Meaning of Intonation Contours

The units such as pitch accents and boundary tones which create intonational contours each have a help and are examined during interpretation of an intonation contour. Tune, namely, intonation contour is determinant of intonation types. The difference among intonation types arises from changes in tune. The choice of a specific tune can be interpreted as the aim of delivering a message to addressee and it is based on both utterance and belief of addressee. Basically, to answer a question falling tune is employed whereas the rising tune carries the meanings such as uncertainty, doubt, and repetition of an answer in a way of question (Pierrehumbert & Hirschberg, 1990).

2.6.1 H* versus L*

The accent type H* is to indicate the utterance is new and tried to be added to the belief of the addressee. With the use of L* pitch accent, the syllable becomes more "salient". The aim is not to add something to the belief between speaker and addressee. If accented tone is L such as L*, L*+H, and H+L* the speaker does not aim to change belief of the addressee and is not certain whereas if accented tone is H such as H*, L+H*, and H*+L the aim of the speaker is to change the belief of the addressee (Pierrehumbert & Hirschberg, 1990, p.292).

2.6.2 H% versus L%

Boundary tones are indicators of "forward-looking". H% boundary tone connects the utterance to the following one. It gives the feeling of incompleteness. HH% (yes-no question) needs to be completed by the following utterance. The turn is changed to the addressee. Hence, replies are given with LL% boundary tone. An utterance with L% bounday tone does not need to be construe with the following utterance. As the falling at the end of an utterance increases, the meaning which is conveyed as completeness becomes more complete. LH% alone does not require a reply; however, it is construe with the following utterance with LL% boundary tone (Pierrehumbert & Hirschberg, 1990, p. 308). The falling boundary tone signals the completeness of the utterance thus of message. On the other hand, rising boundary tone is an indicative of incompleteness and the utterance is followed by another utterance (Ergenç, 2002).

2.6.3 H* and boundary tones

Downstepped contours are as exemplifed with H* LL% intonation contour are indicated as having declarative meaning. Both higher pitch range and low boundary tone are the cues for certainty. Another use of H* LL% intonation contour is wh-questions (Pierrehumbert & Hirschberg, 1990). Moreover, the use of downstepped contours signals certainty regarding epistemic stance of the speaker as indicated in the study by Gravano, Benus, Hirschberg, German, & Ward (2008). Grice & Savinot (1997) found that when the degree of confidence of speaker about information status that is shared with the addressee is low L+H* pitch accent is included in yes-no questions whereas H*+L pitch accent included in the case of high confidence.

H* HL% intonation contour is employed to "elaborate" and "provide support" (Ward & Hirschberg, 1985, p. 291).

H*+L intonation contour is employed in "pedagocical" purposes (Pierrehumbert & Hirschberg, 1990, p. 298).

H*+L HL% intonation contour is employed as "calling contour" (Pierrehumbert & Hirschberg, 1990, p. 299).

H*+L LL% intonation contour is employed when "reading instructions" " (Pierrehumbert & Hirschberg, 1990, p. 299).

L+H* LH% intonation contour is observed when marking "correction" and "contrast" (Pierrehumbert & Hirschberg, 1990, p. 296).

2.6.4 L* and boundary tones

L* HH% intonation contour is indicated as having interogative meaning. To be able to convey the meaning of uncertainty together with when asking yes-no questions, L* HH% intonational contour is employed. This L* gives the possibility in terms of the beliefs of the interlocutors: speakers' belief is incorrect about their belief (Pierrehumbert & Hirschberg, 1990).

L* LH% is employed when either addressee is already sharing the same belief or addressee is expected to share the same belief in deed addressee does not. The latter one shows "insulting effect" (Pierrehumbert & Hirschberg, 1990, p. 292).

Intonation contours such as L*+H LH%, L*+H HH%, and L*+H HL% are described as signaling speaker uncertainty due to L*+H pitch accent based on the work of Ward & Hirschberg (1985).

H+L* LL% is employed when the addressee has knowledge of the utterance with this intonation contour.

2.7 Turkish Intonation

Intonation is not universal. Also, there is a prosodic typology such that some languages are "tonal" while others are not. Among this classification, Turkish is considered as a non-tonal language.

According to Kornfilt (1997) in Turkish, stress and pitch are related. Within a sentence, the constituent before the verb carries the stress thereby in order to find the maximum intonation the stressed constituent should be found. In the same way, Demircan (1983) described intonation as the changes in pitch of voice. Firstly, the focus should be determined to be able to compare pitch contours. In Turkish the focused word can be found before verb. Then, decision of tone can be made.

Intonation contours in Turkish are grouped into three types as "slight rise followed by fall" that is employed in statements, "high rise followed by fall" that is employed in yes-no questions, and "slight rise followed by fall-rise" that is employed in wh-questions, "conditional clause", "adverbial clause", and "Co-ordinated items" (Göksel & Kerslake, 2004, pp. 35-36.).

Falling boundary is observed in question sentences with interrogative particle "-mI (mı, mi, mu, mü)", and in positive and negative declarative sentences. However, in wh- question sentences, rising boundary is observed as well as in question tags such as "isn't it?" 'Question tag' belongs to British English whereas 'tag question' is used in American English. In English, tag questions are employed either to learn the unknown answer or to make sure what is already certain about. An answer seeking tag question demonstrates rising pitch whereas in a case speaker already knows the answer falling pitch is observed. With respect to tag questions uncertainty is signaled with wider pitch range and longer duration (Demirezen, 2014). Moreover, rising pitch which is interrogation intonation is associated with "questions" whereas falling pitch is associated with statements which are "declaration", "approve", and "reply" (Üçok, 1951, p. 135).

A Study by Göksel, Kelepir & Üntak-Tarhan (2007) suggested findings upon intonation in Turkish Question Sentences. In Turkish, the ending intonation is used and it is facilitative when distinguishing a question sentence and a declarative sentence in the existence of ambiguity. This property of intonation reveals the existence of particular pitch contours. The fact remains that the ending is not the only indicative. The example provided in the study illustrated a pattern in which question intonation starts as flat and compressed pitch contours reaching H (high) tone whereas declarative intonation starts with rising-and-falling pitch contours reaching H (high) tone. English intonation differs from that of Turkish since Ipek & Jun (2013) stated that rising pitch is employed to convey certainty.

There are a few studies that examined Turkish intonational phonology by using Autosegmental-Metrical model (see Ipek & Jun, 2013; Özge, 2003; Kamalı, 2011; Kan, 2009). Kan (2009) proposed a model for Turkish intonation and prosody.

However, Ipek & Jun (2013) is the critical of the phonological model suggested by Kan (2009) due to its lack of particularity. The model included pitch accents (i.e., H*, !H*, L+H*, and L+!H*) and boundary tones (i.e., L+H- or L+!H-). After, Kamalı (2011) suggested a model for phonology of Turkish intonation. According to model, in order to have pitch accent utterances should not have final stress. If final syllable has high tone, in other words stress, it is considered as high boundary tone and marked by H tone. Ipek & Jun (2013) conducted a research based on neutral sentences. Length and stress were one of the independent variables manipulated in the study. Characteristic pitch accent of Turkish is H* (Ipek & Jun, 2013). Ipek & Jun (2013) argued that H* pitch accent is considered as being the stressed syllable without paying attention to whether stress is at the final syllable. Ipek & Jun (2013) described L H* tones and L% boundary tones as belong to Turkish declarative sentence. H* pitch accent is observed in stressed syllables.

2.8 Pragmatic Meaning and Demonstratives

"Demonstratives are used to identify a referent in the surroundings of the interlocutors or the addressee's mental/memory representation of a referent" (Piwek, Beun, & Cremers, 2007, p. 4). The use of demonstratives is spatial thus reflects physicality (Küntay & Özyürek, 2006). Cognitive hypothesis presents the dependence between language acquisition and cognitive performance before language acquisiton. There is a link between non-linguistic and linguistic stage. With respect to spatial semantic development, attributes that are learned while discovering objects and their locations in cognitive development, appear in spatial language (Gumperz & Levinson, 1991). Demonstratives are common in all languages and at least two types. Demonstratives are classified based on the relationship between speaker and antecedent. These types are proximal and distal (Diessel, 1999).

In Turkish, there are three demonstrative pronouns which are 'bu' (this), 'şu' (this/that), and 'o' (that). They are employed to indicate nouns and substitute for them. In order to integrate these demonstrative pronouns into a name of place, the suffix –rA- is added. Noun forms are 'bura' (this place), 'şura' (this/that place), and 'ora' (that place). Moreover, noun form can be transformed into locative case. Locative case markers are –de and –da. These two variants are due to vowel harmony (Underhill, 1976, p. 137). Locative forms are 'burada' ((in) here), 'şurada' ((in) here), and 'orada' ((in) there) (Göksel & Kerslake, 2004, p. 47). Since accent is not on the suffix instead it is on the pronoun, the vowel A within the suffix –rA- is not articulated in spoken language (Underhill, 1976). The addition of locative case marker results in the fall of final vowel of name of place. The articulations in speech are 'burda' ((in) here), 'şurda' ((in) here), and 'orda' ((in) there)) (Göksel & Kerslake, 2004, p. 47).

Turkish tripartite demonstrative system is different from that of English where there is only a two-fold distinction between 'here' and 'there'. This is a particularity of Turkish and therefore of special interest. Difference among distances is recognized as an elementary factor to construe semantics of demonstrative system due to spatial

characteristics. For an account based on the difference among distances, categories such as nearby, far, and medial are created based on a centre (Burenhult, 2003). Not only distance-proximity relation but also existence of antecedent within the peripheral perception of addressee is considered when assigning these pronouns to nouns. Kornfilt (1997) has presented a system accounting pronouns according to their distance to speaker. 'Bu' (this) is employed when antecedent is at the closest distance to speaker, for 'şu' (that) it is at the middle distance, and for 'o' (that) it is at the furthest distant. This account differs from that of Lyons' 1977 work (as cited in Küntay & Özyürek, 2006) in terms of including both speaker and addressee. According to this system as in Kornfilt's (1997) system 'bu' (this) is employed when antecedent is at the closest distance to speaker whereas 'şu' (that) is employed if antecedent is at the closest distance to addressee. As to 'o' (that), antecedent is far from both speaker and addressee. However, what these systems present contrast with that of Özyürek (1998) who claimed that the difference between 'bu' (this) and 'o' (that) is likely on the account of distance-proximity relation along with attention of the addressee on the antecedent. However, the system fails to differentiate 'şu' (that). 'Şu' (that) is not employed for an antecedent which is in the middle or at the closest distance to the speaker. The usage of 'şu' (that) is explained by attention of the addressee (i.e. gaze direction) on antecedent. 'Şu' (that) is used to draw attention of the addressee on the antecedent thus pragmatic function is dedicated to this pronoun. Moreover, Küntay & Özyürek (2006) conducted a study with 4 and 6 year old children to examine making use of demonstratives in conversation. Locative, dative, and accusative forms of demonstratives are included in data coding. The results indicated that semantic spatial encoding of "bu" takes into account the distance of the referent thus used for proximity. However, pragmatic function of "şu" emerged and it required visual attention of the addressee on the referent in order to be emloyed. Moreover, 4 year olds were not able to use "şu" whereas for 6 year olds proximity based on the distance-based account surpassed the pragmatic attention drawing function.

There are demonstratives in different languages mimic the same pattern that of Turkish demonstrative system uses via 'şu' (that) in terms of pragmatic function. 'Ton' is one of the demonstratives in Jahai language. Its function is defined as addressee-anchored and accessible. Although it was examined and concluded that it is employed by the speaker when the antecedent is proximal to the addressee, others suggested that encoding in 'ton' does not relate to space. In order to employ 'ton', location of the antecedent is not of the essence in case the antecedent is attended to by the addressee as well. However, this attention means cognitive accessibility to the antecedent rather than having physical features such as proximal, reachable, or visible (Burenhult, 2003). Burenhult (2003) holds the view that pragmatic function of 'şu' (that) is associated with cognitive accessibility as regards both speaker and addressee. In terms of speaker, accessibility comprises "reachability", "approachability", "perceptibility", "possession", etc (p. 365). On the other hand, regarding addressee, accessibility comprises "attention" or "knowledge" (p. 366). Cognitive relation between antecedent and addressee are distinguished. This pragmatic function overrides distance-proximity account. Likewise, preference for

demonstrative pronouns with respect to attention of the addressee was observed in the languages Tiriyo and Brazilian Portuguese. For the systems with distance-proximity account due to the pragmatic importance of addressee semantic meaning of the pronoun has attached. Center of the system has shifted from distance to addressee (Meira, 2003). A counter examle from Dutch language suggested an example for how proximity or distance of a referent based on speaker fails. In this example dialogue, the patient used the distal type while mentioning on his body part whereas the doctor used proximal type for the same antecedent. In this example, addressee was the baseline (Janssen, 1995).

A corpus study by Piwek, Beun, & Cremers (2007) in Dutch language offered findings on the use of demonstratives with respect to distance-proximity relation and compared them to English. In English two types are; "this/these" and "that/those", respectively. Dutch variants are "dit/deze" and "die/dat". The data was elicited from speech and consisted of references to objects in a setting among interlocutors. In addition to general account of distance-proximity relationship, the writers suggested a new perspective and term "indicating" which means attention drawing to antecedent (Piwek, Beun, & Cremers, 2007, p. 3). Drawing attention can be made by pitch thus level of intensity is manipulated. They hypothesized that levels of indicating are intense and neutral in which they are matched with proximals and distals, respectively. Moreover, hypothesis included low accessibility and high importance to antecedent. Accessibility was defined as "the ease (of effort) with which particular mental contents come to mind" (Kahneman, 2003, p. 699). In the study authors accepted the definition "focus of attention" (Piwek, Beun, & Cremers, 2007, p. 1). If something is distant, it has low accessibility. Speakers would prefer proximals for low accessibility and more importance whereas they would prefer distals for high accessibility and less importance. The results showed a relation between intense indicating and low accessibility, meanly using proximals for low accessibility. However, the results showed no relationship between intense indicating and high importance, meanly using proximals for high importance. Piwek, Beun, & Cremers (2007) in their study provided a cognitive model for deciding on demonstratives among distal and proximal ones to indicate accessibility and importance.

2.9 Pragmatic Meaning and Modality

Stance of speakers can be expressed by using modals. Conveyance of stance has pragmatic aspect from the point of addressee. Understanding stance of speaker facilitates grasping the meaning of utterance (Green, 1979). Utterances bring along a specific confidence level. Our feeling of knowing about what we uttered might be in different confidence levels. In pursuance of our confidence we use lexical items such as adverbs and manipulate intonation (Dral, Heylen, op den Akker, 2011).

Epistemic adverbs also known as adverbials function fundametally in expressing modality. They help speakers express their stance by creating both trust and

incredulity. These meanings are grasped by the addressees due to the semantics of adverbials.

Moore, Harris, & Patriquin (1993) observed how children would take notice of intonation and belief words while evaluating speaker confidence in an experiment that required a guess about a hidden object under one of the two boxes. Two puppets gave clues regarding the location of the object. Stimuli were contradicting utterances of two puppets. The children were between 3-6 years old. An English belief term (i.e., know, think, or guess) matched with one of two types of pitch contour at the end of utterance (i.e., high or low). According to results children started to use intonation as a cue at the age of four and found low intonation more trustworthy. As the age of children increased, they were more likely to use belief terms as cues without noticing intonation. The stimuli in the second experiment comprised utterances that were both congruent and incongruent pairings of belief words and types of intonation (i.e., certain belief word was a match for high intonation). According to results intonation was not a usable cue for four year olds to decide on the location of the hidden object based on the utterances of puppets. Only belief words helped children to decide between two locations. However, for five year olds intonation gained prominence. To conclude from a developmental point of view children initially take notice of lexical information. Later on, prosodic information is added to lexical information thus they function together pragmatically. The authors specified the need of pragmatic context to find out the prominence of intonation in children's comprehension. Likewise, Moore, Pure, & Furrow (1990) grounded their experiments based on theory of mind. For the first experiment the task was the same with the work of Moore, Harris & Patriquin (1993) except for lexical stimuli were modal verbs (i.e., must, might, or could) and modal adjuncts (i.e., probably, possibly, or maybe) of all pairings. The second experiment included mental terms in addition to modals. It is concluded that increasing success at later ages is due to the development of false belief at the age of four. False belief helped children in comprehension of speaker confidence.

2.10 Pragmatic Meaning and Confidence Studies

Listeners infer confidence level of the speakers by using different cues. Thus, Comprehension of speaker confidence leads to more practical communication. There are definite nonverbal cues that are used by speakers in production of certainty and uncertainty, and are attended to by listeners to perceive speaker confidence. This shows significance of nonverbal cues in perception of both certainty and uncertainty. However, it is not well known that which of these cues features in perception of speaker confidence.

Hübscher, Esteve-Gibert, Igualada, & Prieto (2016) studied with children on resolving speaker uncertainty. In the study, they tried to figure out the most reliable one among the cues which are intonational, lexical (i.e., think and maybe), and gestural. Younger children were more likely to rely on gestures and intonation rather than lexicals. It is known that intonation is more reliable for younger children than

lexicals. However, in a previous study it was found that having more than one modality such as both auditory and visual facilitates perception of the level of speaker confidence (Swerts & Krahmer, 2005). Intonation carries information about motor behavior and mental state of the speakers compatible with facial and hand gestures (Demirezen, 2014). Gesture and intonation are in harmony. For instance, falling intonation comes along with downward or forward hand movements whereas rising intonation comes along with upward or backward hand movements. This suggests that pragmatic function of intonation is carried as well by gestures (Bolinger, 1983). Beat gestures proceeds turn taking in a dialogue (Balog & Brentari, 2008).

Numerous studies revealed that if there is incompatibility between verbal and non-verbal communication children make much of lexical meaning rather than auditory meaning in an affective communication by the age of 4 (Friend, 2003). A study on social-referencing by Lawrence & Fernald (1993) compared 9 month olds and 18 month olds (as cited in Friend, 2001). The study suggested that in case of incompatibility between paralinguistic and linguistic content, behaviors of older children monitored by lexicals which is linguistic content. Supportively, in a later research with 15 month olds, it was observed that if the children do not know the meaning of the lexical, paralanguage monitors their behavior (Friend, 2001).

Friend (2001) in their study with infants used both compatible and incompatible linguistic and paralinguistic pairings. Within the consistent condition lexical content and paralanguage were matching while within discrepant condition these two were not congruent. Measured behavior was the time spent with a novel toy. It was expected that paralanguage would monitor the behavior of children in case of incompatibility and for children in transition phase to adultlike pattern lexical meaning would monitor their behavior based on their knowledge of the meaning of lexicals. 15-16 month olds were choosen as the ones in transition phase. The results indicated that paralanguage surpassed linguistic content for behavior regulation at 15 months of age (see Mumme et al. 1996, Sorce et al. 1985, as cited in Friend, 2001). Also, in line with the expectation understanding the lexical meaning, in other words receptive lexicon, provided language for behavior regulation. In such a case language would have influence on infant behavior.

“The melody carries the message in speech addressed to infants to a much greater extent than in speech addressed to adults” (Fernald, 1989, p. 1505). Krahmer & Swerts (2005) conducted a production and a perception study with Dutch participants. Production study was composed of elicitation of certain and uncertain utterances from children and adults. Utterances were elicited from answers to factual questions based upon Feeling of Knowing paradigm (Hart, 1965). According to results of the production experiment, it was revealed that when uncertainty was conveyed; fillers, delays, high intonation, eyebrow movements were employed by both children and adult speakers with small differences. Moreover, perception study included these elicited utterances for an evaluation with respect to uncertainty by both children and adults. The results indicated that correct estimation for uncertainty

was greater for both children and adults while evaluating adult speakers. On the other hand, adults were more successful than children on the use of audiovisual cues. Since, in terms of cognitive development children fall behind adults to grasp speaker confidence. Children reach this ability at the age of four (Moore, Pure, & Furrow, 1990). Among the cues fillers are used only by adults to differentiate uncertainty. Children only expressed uncertainty by using delays due to longer search in the memory; however, they did not employed fillers during delays. In general uncertainty was found to be associated with delays, fillers, and high intonation.

Jiang & Pell (2017) presented in their study that feeling of knowing of the speakers is conveyed both linguistically and vocally thus listeners are able to categorize certainty and doubt which are cognitive states of the speakers. Evaluation of the speaker confidence is regulated by the utterance due to aforethought confidence (i.e., confident, unconfident, and neutral) and its communicative function. Also, evaluation of speaker confidence was regulated by an utterance which indicates probability. According to perception of the addressees, confidence was associated with highest F_0 range whereas doubt was associated with highest mean F_0 and slowest speaking rate. Utterances with most pitch changes and faster speech rate were perceived as confident. Moreover, neutral utterances had lower pitch and pitch range and faster speech rate when compared to confident utterances. Furthermore, utterances with highest pitch, longer duration, and slower speech rate were perceived as unconfident.

Dijkstra, Krahmer, & Swerts (2006) conducted a perception study with adult Dutch speakers to find out whether cues such as fillers, rising intonation, and facial expressions would differentiate from each other with respect to their effect on the perception of speaker confidence. Stimuli included both absence and presence of filler. Moreover, both falling and rising intonation, and both neutral and marked facial expressions. The combinations were incongruent. Participants evaluated certainty level of utterances which were manipulated answers to factual questions. Results indicated that all these cues have effect on the perception of certainty but the effect of facial expressions outstood in which a marked expression conveyed uncertainty. With respect to intonation production of an utterance with a rising intonation decreased perception of certainty. The effect of fillers was found at the least.

According to Dijkstra, Krahmer, & Swerts (2006), high feeling of knowing (certain) was associated with absence of fillers whereas low feeling of knowing (uncertain) was associated with presence of fillers. Moreover, falling intonation was considered as a signal for certainty whereas rising intonation was considered as a signal for uncertainty. Furthermore, a neutral face indicated high feeling of knowing while marked facial expression was indicating uncertainty (p. 1). Jokinen (2010) and Pon-Barry & Shieber (2010) described unconfidence by high pitch and slower speech rate (as cited in Jiang & Pell, 2017). The strategies of conveying uncertainty while speakers are unsure of their answers to questions includes use of fillers and rise in the intonation (Dijkstra, Krahmer, & Swerts, 2006). Also, intonation contour can be

in question form (Smith & Clark, 1993). Uncertainty is conveyed by the choice of marked cues whereas neutral cues are preffered to express certainty (Dijkstra, Krahmer, & Swerts, 2006). Dral, Heylen, & op den Akker (2011) examined prosodic properties in speech and wanted to bring on an automatic way of recognizing (un)certainty in speech. The study aimed to determine confidence of adult speakers about correctness of utterances they uttered. They presented that low speed of talking, low intensity, and rising pitch as properties of uncertainty.

The next two chapters continue with the methods which were employed in production and perception experiments such as the task for utterance elicitation, location, materials, procedure, and recordings of elicited responses. Mooreover, the analysis, results, and short discussions will be provided.

CHAPTER 3

EXPERIMENT 1

3.1 Aim

The production experiment aimed to create a set of utterances that conveyed speaker confidence (i.e., certain, uncertain, and neutral) and collection of utterances to conduct perception experiment to detect epistemic confidence levels together with the specification of acoustic properties and lexical markers that were characteristic of epistemic confidence perception in communication.

There are definite nonverbal cues that are used by speakers in production of certainty and uncertainty. The production study concentrated on confidence of the speakers (i.e., certain and uncertain) to make an observation and have an understanding on the use of acoustic properties. The main question of this study is to determine how epistemic stance is conveyed by means of verbal (i.e., epistemic adverbs and locative pronouns) and acoustic cues (i.e., pitch, pitch range, and duration).

How acoustic properties are manipulated and preferred to express the epistemic confidence by speakers was the question; therefore, the present experiment was conducted in order to elicit utterances expressing "certainty" and "uncertainty" of the speaker. In accordance with this purpose, participants voiced pronouns and adverbs given to them as well as their combinations, within an imaginary scenario. While utterances were elicited speakers were required to answer in a natural way toward to their epistemic stance and the experimenter did not exemplify regarding how utterances were supposed to be voiced (Jiang & Pell, 2017). Audio recordings were made while the participants uttered these word classes. The collected auditory materials were analyzed to get an idea about the prosodic strategies that participants used when expressing certainty and uncertainty.

3.2 Participants

The present experiment included audio recordings of fifteen female participants. However, after eliminating unusable audio recordings, the set of audio clips used for later analysis belonged to 7 female native speakers of Turkish (M = 29.142, SD = 7,081). Humming, stuttering, tongue slips, artificial voice, rustling, and/or lack of strategy exemplify reasons to eliminate participant data. The reason for involving only female participants is that female voice has higher pitch range than male voice

and it is easier to track comparisons. Voluntary participants were recruited by using the snowball sampling method. Participants consisted of Master and PhD students of Middle East Technical University (METU), Ankara.

3.3 Materials & Design

The audio recording sessions took place in Cogs Lab at the Informatics Institute of METU. The software version 2.1.3 of Audacity ® (Audacity Team, 2017)[2] was used to record the sessions and run on an Intel® Core™ i7 Lenovo-PC running Windows 8.1 64-bit with Conexant SmartAudio HD driver and a refresh rate of 60p Hz. Earphones with microphone were used as recording device. Stimuli in the experiment consisted of an imaginary scenario and word classes to be used in these scenarios. The imaginary scenario enabled participants to utter the word classes provided in Tables 3.1-3.3. Word classes were composed of locative pronouns, namely "burada" (here), "orada" (there), "şurada" (there); epistemic adverbs expressing certainty and uncertainty, namely "kesinlikle" (certainly) and "galiba" (probably), respectively; and locative pronouns modified by epistemic adverbs expressing certainty and uncertainty, namely "kesinlikle burada" (it is certainly here), "kesinlikle orada" (it is certainly there), "kesinlikle şurada" (it is certainly there), "galiba burada" (it is probably here), "galiba orada" (it is probably there), "galiba şurada" (it is probably there).

Table 3.1: Locative pronouns

Means of expression	**Modality**	**Locative pronoun**	**Requested utterance**
Locative pronoun	certain	burada	Burada [certain stance]
		orada	Orada [certain stance]
		şurada	Şurada [certain stance]
	uncertain	burada	Burada [uncertain stance]
		orada	Orada [uncertain stance]
		şurada	Şurada [uncertain stance]

[2]Audacity® software is copyright © 1999-2018 Audacity Team.

The name Audacity® is a registered trademark of Dominic Mazzoni.

Table 3.1 illustrates the locative pronouns, namely “burada” (here), “orada” (there), “şurada” (there). They were instructed to utter the stimuli in a “certain” or “uncertain” stance. Participants took both certain and uncertain stance while uttering them. Each speaker voiced mentioned locative pronouns once and took both certain and uncertain stance. In total, materials included 42 audio clips belonging to 7 speakers.

Table 3.2: Lexical words

Means of expression	**Modality**	**Lexical word**	**Requested utterance**
Lexical word	certain	kesinlikle	Kesinlikle [certain stance]
	uncertain	galiba	Galiba [uncertain stance]
	neutral	kesinlikle	Kesinlikle [neutral stance]
	neutral	galiba	Galiba [neutral stance]

Table 3.2 presents the epistemic adverbs, namely “kesinlikle” (certainly) and “galiba” (probably). Participants took both certain and neutral stance while uttering “kesinlikle” (certainly). Moreover, participants took both uncertain and neutral stance while uttering “galiba” (probably). Each speaker voiced mentioned epistemic adverbs once. In total, materials included 28 audio clips belonging to 7 speakers.

Table 3.3: [lexical word] + [locative pronoun]

Means of expression	Modality	Lexical word	Locative pronoun	Requested utterance
Lexical word + Locative pronoun	certain	kesinlikle	burada	Kesinlikle [certain stance] + burada [neutral stance]
			orada	Kesinlikle [certain stance] + orada [neutral stance]
			şurada	Kesinlikle [certain stance] + şurada [neutral stance]
	uncertain	galiba	burada	Galiba [uncertain stance] + burada [neutral stance]
			orada	Galiba [uncertain stance] + orada [neutral stance]
			şurada	Galiba [uncertain stance] + şurada [neutral stance]

Table 3.3 exhibits locative pronouns modified by epistemic adverbs, namely "kesinlikle burada" (it is certainly here), "kesinlikle orada" (it is certainly there), "kesinlikle şurada" (it is certainly there), "galiba burada" (it is probably here), "galiba orada" (it is probably there), "galiba şurada" (it is probably there). Participants took certain stance while uttering epistemic adverb, namely "kesinlikle" (certainly) and neutral stance while uttering locative pronouns, namely "burada" (here), "orada" (there), "şurada" (there). Moreover, they took uncertain stance while uttering epistemic adverb, namely "galiba" (probably) and neutral stance while uttering locative pronouns, namely "burada" (here), "orada" (there), "şurada" (there). Each speaker voiced mentioned utterances once. In total, materials included 21 audio clips belonging to 7 speakers.

3.4 Procedure

The study was approved by METU ethics committee (see Appendix A). The experimenter ran the software version 2.1.3 of Audacity ® (Audacity Team, 2017) on a Lenovo-PC. Earphones with microphone were plugged in and checked. The preferred settings for the audio host and recording channels were left at their default settings. The participants were invited into the Cogs Lab and seated at a table with the PC. After explaining the purpose and the ethical concerns of the experiment they confirmed the consent form to make sure they could attend the experiment (see Appendix B). The instructions were given by the experimenter verbally. According

to the imaginary scenario, a key was hidden. According to the imaginary scenario the participants were required to imagine the experimenter entered the room asking "Where is the key?" and answer to that question. The participants should imagine the object which was a hidden key somewhere. They were asked to take a certain/uncertain/neutral stance. They were informed about the steps of the recording session (see Appendix C). The participants were presented to the same material in the same order, i.e., a within-subjects design was employed.

The instructions they were given only included the various scenarios and word classes as stimuli. How participants would express certainty and uncertainty in different ways, prosodically, was left to them. They were encouraged to produce the requested utterances in a way which is natural. The experimenter did not provide an example of how the requested utterances would be expressed.

This procedure enabled each participant to develop their own strategy when uttering the requested words or word combinations. The sessions took 15 minutes to complete, on average. Audio recordings were exported as wav files in 32-bit float format with a sample rate of 44100 Hz. The experimental session terminated with the debriefing process in which the participants were handed over the debriefing form (see Appendix D). After all experiments were completed, the experimenter listened to the audio files and eliminated some of due to the criteria mentioned above. The remaining data files belonging to 7 female native speakers of Turkish were taken into account for the subsequent analysis. The audio recordings had been made continuously. For that reason, audio files were cut into parts to obtain a single audio clip for each utterance. The audio files in wav format were opened by the software version 2.1.3 of Audacity ® (Audacity Team, 2017). Based on the wave form, the regions with an audio signal were selected by moving the cursor to these regions. These selections were listened to and thus the regions with the auditory signals were detected. The selected region was exported in order to obtain each utterance as a single file in wav format. Each speaker voiced sixteen utterances. In total, 112 audio clips were obtained belonging to the 7 speakers. These were created in order to be used in Experiment 2 and in the intonation analysis part.

3.5 Data preparation

Data preparation for the analysis of acoustic properties was conducted by Praat software (Boersma & Weenink, 2017). The audio files in wav format were opened in the software as objects by "Read from file" command. For each wav file "View & Edit" command was used to display sound and in the opened window based on the wave form, the regions with an audio signal were selected by moving the cursor to these regions. These selections were listened to and thus the regions with audio were detected. The selected area was extracted as a single sound in wav format. Thus, sound trimming was completed. The trimmed wav files were opened again in the software as objects by "Read from file" command. For each wav file "View & Edit" command was used to display sound and in the opened window all regions were selected by moving cursor. The following acoustic properties are collected manually: "get pitch" command yielded mean pitch in selection, "get maximum pitch"

command yielded maximum pitch in selection, and "get minimum pitch" command yielded minimum pitch in selection in Hertz. For the purpose of calculating pitch range, the difference between maximum and minimum pitch in selection was taken into account. "Get total duration" command yielded the duration of the recording in seconds. This acoustic property was converted to milliseconds.

3.6 Block I

When the participants imagined to know the place of the key and were sure of it, they were required to indicate the location of the key by using only locative pronouns, i.e., "burada" (here), "orada" (there), and "şurada" (there) in a "certain" way. The scenario and the requirement were repeated for each locative pronoun by the experimenter. When they imagined not to know the place of the key and were unsure of it, they were required to indicate the location of the key by using those pronouns as well, but now in "uncertain" way. The same procedure was applied (for this condition, see Table 3.1). This material was prepared in order to use in Block I as stimuli.

3.6.1 Result

A One-way repeated measures ANOVA was ran to compare the effect of modality (certain, uncertain) on mean pitch. The main effect of modality yielded an F ratio of $F(1,20) = 1.206$, $p = 0.285$, partial $\eta 2 = .057$, indicating that mean pitch did not have statistically significant effect on modality. However, mean pitch was higher for certain modality ($M = 225.29$, $SD = 28.35$) than for uncertain modality ($M = 215.42$, $SD = 20.90$). Mean pitch of the items categorized as certain and uncertain were so close ($M_c = 220.46$ and $M_u = 220.48$, respectively).

A One-way repeated measures ANOVA was ran to compare the effect of modality on duration. The main effect of modality yielded an F ratio of $F(1,20) = 122.837$, $p < .666$, partial $\eta 2 = .010$, indicating that modality did not yield statistically significant differences in duration. Although the result was not statistically significant, duration was longer for uncertain modality ($M = 585.37$, $SD = 124.009$) than for certain modality ($M = 502.95$, $SD = 82.358$).

A One-way repeated measures ANOVA was ran to compare the effect of modality on pitch range. The main effect of modality yielded an F ratio of $F(1,20) = 0.478$, $p = 0.497$, partial $\eta 2 = .023$, indicating that modality did not have statistically significant effect on pitch range. However, mean pitch was higher for certain modality ($M = 146.42$, $SD = 76.06$) than for uncertain modality ($M = 134.66$, $SD = 59.34$).

A One-way repeated measures ANOVA was ran to compare the effect of locative on duration. The main effect of locative yielded an F ratio of $F(2,26) = 27.264$, $p < 0.001$, partial $\eta 2 = .677$. Bonferroni post hoc test indicated that duration was significantly shorter for locative "burada" (here) (486.14 ± 89.35 msec, $p < .001$) and for locative "orada" (there) (508.37 ± 75.12 msec, $p < .001$) compared to locative "şurada" (there) (637.98 ± 107.09 msec). There was no statistically

significant difference between locative “burada” (here) and locative “orada” (there) (p = 0.529).

A two-way ANOVA was conducted that examined the effect of locative pronoun and modality on duration. There was no statistically significant interaction between the effects of locative pronoun and modality for duration, $F(2, 36) = 1.212$, $p = .309$, partial η2 = .063. The estimated marginal means of durations for certain “burada” (here) and uncertain “burada” (here) were 459.313 (*SE* = 30.789) and 512.964 (*SE* = 30.789), respectively. The estimated marginal means of durations for certain “şurada” (there) and uncertain “şurada” (there) were 569.099 (*SE* = 30.789) and 706.858 (*SE* = 30.789), respectively. The estimated marginal means of durations for certain “orada” (here) and uncertain “orada” (here) were 480.441 (*SE* = 30.789) and 536.301 (*SE* = 30.789), respectively.

3.6.2 Discussion

We should note that choice of parameters which are mean pitch, duration, and pitch range did not differ significantly among modalities (i.e., certain and uncertain). However, there found to be some tendencies. Certain modality was produced with higher pitch than uncertain modality. High pitch was choosen to indicate certainty. Moreover, uncertain modality was produced longer than certain modality. This conflicts with the finding with respect to mean F0, the utterances which were evaluated as unconfident had highest value (Jiang & Pell, 2017). High and/or rising fundamental frequency conveys politeness and lack of confidence while low and/or falling fundamental frequency conveys aggression and confidence (Bolinger, 1978; Ohala, 1984). However, English intonation differs from that of Turkish since Ipek & Jun (2013) stated that rising pitch is employed to convey certainty. Longer duration was chosen to indicate uncertainty. With respect to speaking rate, the utterances which were evaluated as unconfident had slowest rate. Utterances with highest pitch, longer duration, and slower speech rate were perceived as unconfident (Jiang & Pell, 2017). Furthermore, certain modality was produced with higher pitch range than uncertain modality. High pitch range was chosen to indicate certainty. The dynamic change of pitch helped to distinguish between confidence and unconfidence. With respect to fundamental frequency range, the utterances which were evaluated as confident had highest value (Jiang & Pell, 2017). Among acoustic features to indicate certainty, only mean pitch value used different from the English literature. However, as in previous studies longer duration is used to indicate uncertainy whereas higher pitch range is used for certainty. In addition, the effect of the locative pronoun was found on duration. The locative “şurada” (there) uttered in significantly longer duration than “burada” and “orada”. The locative “burada” (here) had the shortest duration. It can be suggested that the reason for uttering “burada” (here) in shortest way might be due to the need for emphsizing on certainty. However, this effect might be due to the locative pronoun itself is shorter. With respect to duration of an utterance to be able to suggest that value of an utterance is longer or shorter there is need for comparisons between mean and obtained values (Dral, Heylen, & op den Akker, 2011).

3.7 Block II

The participants were required to utter only the adverb “kesinlikle” (certainly) with a definite modality in the context of the scenarios explained above. They were given the information that they should use the adverb “kesinlikle” (certainly) for taking a certain stance and they were requested to utter “kesinlikle” (certainly) in certain way. In addition, they were requested to utter the adverb “kesinlikle” (certainly) in neutral way, in order to have a prosodic baseline for this epistemic adverb (for this condition, see Table 3.2).

3.7.1 Result

A One-way repeated measures ANOVA was ran to compare the effect of modality on mean pitch. The main effect of modality yielded an F ratio of $F(1,6) = 0.247$, $p = 0.835$, partial $\eta 2 = 0.008$, indicating that modality did not have statistically significant effect on mean pitch. However, mean pitch was higher for certain modality ($M = 219.88$, $SD = 28.898$) than it is for neutral modality ($M = 217.04$, $SD = 18.023$).

A One-way repeated measures ANOVA was ran to compare the effect of modality on duration. The main effect of modality yielded an F ratio of $F(1,6) = 2.631$, $p = 0.156$, partial $\eta 2 = 0.305$, indicating that modality did not have statistically significant effect on duration. However, duration was longer for neutral modality ($M = 802.19$, $SD = 109.844$) than it is for certain modality ($M = 732.69$, $SD = 101.935$).

A One-way repeated measures ANOVA was ran to compare the effect of modality on pitch range. The main effect of modality yielded an F ratio of $F(1,6) = 0.213$, $p = 0.661$, partial $\eta 2 = 0.034$, indicating that modality did not have statistically significant effect on pitch range. However, pitch range was higher for certain modality ($M = 116.57$, $SD = 61.26$) than it is for neutral modality ($M = 103.00$, $SD = 57.869$).

3.7.2 Discussion

We should note that choice of parameters which are mean pitch, duration, and pitch range did not differ significantly among modalities (i.e., certain and neutral). However, there found to be some tendencies. Certain modality was produced with higher pitch than neutral modality. High pitch was choosen to indicate certainty. This result is different than the previous literature since it was found that with respect to mean F0, the utterances which were evaluated as unconfident had highest value (Jiang & Pell, 2017). Moreover, rising intonation and lack of confidence are adapted (Smith & Clark, 1993). On the other hand, English intonation differs from that of Turkish since Ipek & Jun (2013) stated that rising pitch is employed to convey certainty. However, the result is in line with the finding that neutral stance is conveyed by medium pitch level (Rodero, 2011). Moreover, neutral modality was produced longer than certain modality. Shorter duration was chosen to indicate certainty. Dral, Heylen, & op den Akker (2011) suggested that low speed of talking,

low intensity, and rising pitch as properties of uncertainty. Furthermore, certain modality was produced with higher pitch range than neutral modality. The dynamic change of pitch helped to distinguish between confidence and unconfidence. With respect to fundamental frequency range, the utterances which were evaluated as confident had highest value (Jiang & Pell, 2017). High pitch range was chosen to indicate certainty. Among acoustic features to indicate certainty, only mean pitch value used different from the English literature. However, as in previous studies longer duration is used to indicate uncertainy whereas higher pitch range is used for certainty.

3.8 Block III

The participants were required to utter only the adverb "galiba" (probably) with a definite modality in the context of the scenarios explained above. They were given the information that they should use the adverb "galiba" (probably) for taking an uncertain stance and they were requested to utter "galiba" (probably) in uncertain way. In addition, they were requested to utter the adverb "galiba" (probably) in neutral way, in order to have a prosodic baseline for this epistemic adverb (for this condition, see Table 3.2).

3.8.1 Result

A One-way repeated measures ANOVA was ran to compare the effect of modality on mean pitch. The main effect of modality yielded an F ratio of $F(1,6) = 0.28$, $p = 0.873$, partial $\eta 2 = 0.005$, indicating that modality did not have statistically significant effect on mean pitch. However, mean pitch was higher for neutral modality ($M = 215.54$, $SD = 36.368$) than it is for uncertain modality ($M = 213.16$, $SD = 26.726$).

A One-way repeated measures ANOVA was ran to compare the effect of modality on duration. The main effect of modality yielded an F ratio of $F(1,6) = 3.583$, $p = 0.107$, partial $\eta 2 = 0.374$, indicating that modality did not have statistically significant effect on duration. However, duration was longer for uncertain modality ($M = 661.62$, $SD = 122.214$) than it is for neutral modality ($M = 577.30$, $SD = 90.956$).

A One-way repeated measures ANOVA was ran to compare the effect of modality on pitch range. The main effect of modality yielded an F ratio of $F(1,6) = 0.185$, $p = 0.682$, partial $\eta 2 = 0.030$, indicating that modality did not have statistically significant effect on pitch range. However, pitch range was higher for uncertain modality ($M = 141.42$, $SD = 52.45$) than it is for neutral modality ($M = 127.77$, $SD = 76.178$).

3.8.2 Discussion

We should note that choice of parameters which are mean pitch, duration, and pitch range did not differ significantly among modalities (i.e., uncertain and neutral).

However, there found to be some tendencies. Neutral modality was produced with higher pitch than uncertain modality. Supporting that neutral stance is conveyed by medium pitch level (Rodero, 2011). The result also support that English intonation differs from that of Turkish since Ipek & Jun (2013) stated that rising pitch is employed to convey certainty. Higher pitch was choosen to indicate certainty. However, it is in conflict with the findings that are with respect to mean $F_{0,}$ the utterances which were evaluated as unconfident had highest value (Jiang & Pell, 2017) and rising intonation and lack of confidence are adapted (Smith & Clark, 1993). Moreover, uncertain modality was produced longer than neutral modality. Shorter duration was chosen to indicate certainty. With respect to speaking rate, the utterances which were evaluated as unconfident had slowest rate (Jiang & Pell, 2017). Utterances with highest pitch, longer duration, and slower speech rate were perceived as unconfident (Jiang & Pell, 2017). Dral, Heylen, & op den Akker (2011) suggested that low speed of talking, low intensity, and rising pitch as properties of uncertainty. Furthermore, uncertain modality was produced with higher pitch range than neutral modality. High pitch range was chosen to indicate uncertainty. The result is not in line with the previous findings. With respect to fundamental frequency range, the utterances which were evaluated as confident had highest value (Jiang & Pell, 2017). The results are similar to previous literature except that the manipulation of pitch range. Neutral modality was expected to have wider pitch range.

3.9 Block IV.I

The combinations of the adverb “kesinlikle” (certainly) with locative pronouns were required in the last step. The adverb “kesinlikle” (certainly) and the locative pronouns “burada” (here), “orada” (there), and “şurada” (there) were combined one by one. The scenarios and the requests were repeated for every combinations. The participants were requested to use certain stance while uttering “kesinlikle” (certainly), whereas they were requested to utter the locative pronouns in neutral way. Uttered combinations were thus “kesinlikle burada” (It is certainly here), “kesinlikle şurada” (It is certainly there), “kesinlikle orada” (It is certainly there) (for this condition, see Table 3.3).

3.9.1 Result

A One-way repeated measures ANOVA was ran to compare the effect of locative on mean pitch. The main effect of locative yielded an F ratio of $F(2,12) = 0.543$, $p = 0.595$, partial $\eta 2 = 0.083$, indicating that locative did not have statistically significant effect on pitch. However, mean pitch was the highest for “kesinlikle şurada” (It is certainly there) *(*$M = 243.11$, $SD = 24.600$, followed by “kesinlikle burada” (It is certainly here) ($M = 238.96$, $SD = 20.366$) as second, and “kesinlikle orada” (It is certainly there) ($M = 234.86$, $SD = 21.337$).

A One-way repeated measures ANOVA was ran to compare the effect of locative on duration. The main effect of locative yielded an F ratio of $F(2,12) = 2.114$, $p = 0.164$,

partial η2 = 0.260, indicating that locative did not have statistically significant effect on duration. However, duration was the highest for “kesinlikle şurada” (It is certainly there) *(M* = 1498.740, *SD* = 419.275, followed by “kesinlikle burada” (It is certainly here) (*M* = 1388.367, *SD* = 253.602) as second, and “kesinlikle orada” (It is certainly there) (*M* = 1318.128, *SD* = 262.907).

A One-way repeated measures ANOVA was ran to compare the effect of locative on pitch range. The main effect of modality yielded an F ratio of *F*(2,12) = 0.321, *p* = 0.731, partial η2 = .051, indicating that locative did not have statistically significant effect on pitch range. However, pitch range was the highest for “kesinlikle şurada” (It is certainly there) *(M* = 188.64, *SD* = 76.671, followed by “kesinlikle orada” (It is certainly there) (*M* = 179.27, *SD* = 71.091) as second, and “kesinlikle burada” (It is certainly here) (*M* = 169.23, *SD* = 37.749).

3.9.2 Discussion

We should note that choice of parameters which are mean pitch, duration, and pitch range did not differ significantly among locatives. However, there found to be some tendencies. The locative “şurada” combined with the adverb “kesinlikle” had highest pitch, duration, and pitch range. We observed that locative “şurada” had longest duration. This finding is also supportive for our claim that this effect might be due to the locative pronoun itself is longer. With respect to duration of an utterance to be able to suggest that value of an utterance is longer or shorter there is need for comparisons between mean and obtained values (Dral, Heylen, & op den Akker, 2011). We do not have enough evidence to suggest that “kesinlikle şurada” is the most certain utterance according to manipulation of pitch level and pitch range. There needs to be further investigation.

3.10 Block IV.II

The combinations of the adverb “galiba” (probably) with locative pronouns were required in the last step. The adverb “galiba” (probably) and the locative pronouns “burada” (here), “orada” (there), and “şurada” (there) were combined one by one. The scenarios and the requests were repeated for every combination. Uncertain stance was required while uttering “galiba” (probably). However, locative pronouns were required to be uttered in neutral way. Uttered combinations were “galiba burada” (It is probably here), “galiba şurada” (It is probably there), and “galiba orada” (It is probably there) (for this condition, see Table 3.3).

3.10.1 Result

A One-way repeated measures ANOVA was ran to compare the effect of locative on mean pitch. The main effect of locative yielded an F ratio of *F*(2,12) = 1.720, *p* = 0.220, partial η2 = 0.223, indicating that locative did not have statistically significant effect on pitch. However, mean pitch was the highest for “galiba şurada” (It is probably there) *(M* = 240.93, *SD* = 26.660, followed by “galiba burada” (It is

probably here) (*M* = 230.72, *SD* = 29.223) as second, and "galiba orada" (It is probably there) (*M* = 228.95, *SD* = 23.246).

A One-way repeated measures ANOVA was ran to compare the effect of locative on duration. The main effect of locative yielded an F ratio of $F(2,12) = 0.251$, $p = 0.782$, partial $\eta 2 = 0.040$, indicating that locative did not have statistically significant effect on duration. However, duration was the highest for "galiba burada" (It is probably here) *(M* = 1381.396, *SD* = 272.095, followed by "galiba şurada" (It is probably there) (*M* = 1380.547, *SD* = 251.962) as second, and "galiba orada" (It is probably there) (*M* = 1340.975, *SD* = 220.322).

A One-way repeated measures ANOVA was ran to compare the effect of locative on pitch range. The main effect of modality yielded an F ratio of $F(2,12) = 0.566$, $p = 0.582$, partial $\eta 2 = .086$, indicating that locative did not have statistically significant effect on pitch range. However, pitch range was the highest for "galiba orada" (It is probably there) *(M* = 183.66, *SD* = 49.044, followed by "galiba burada" (It is probably here) (*M* = 157.49, *SD* = 73.867) as second, and "galiba şurada" (It is probably there) (*M* = 153.13, *SD* = 65.571).

3.10.2 Discussion

We should note that choice of parameters which are mean pitch, duration, and pitch range did not differ significantly among locatives. However, there found to be some tendencies. The parameters; pitch, duration, and pitch range did not point to a mutual utterance. Highest pitch was found in "galiba şurada" (It is probably there), longest duration was found in "galiba burada" (It is probably here), and highest pitch range was found in "galiba orada" (It is probably there). We do not have enough evidence to conclude that due to highest pitch "galiba şurada" is the least confident utterance. Moreover, we cannot conclude that "galiba burada" is the least confident utterance due to longest duration. Furthermore, highest pitch range belongs to "galiba orada" does not put it in the category of certainty. We need further investigations.

3.11 Summary of the results

High pitch was choosen to indicate certainty. Lower pitch was choosen to differentiate neutrality than uncertainty. Longer duration was chosen to indicate uncertainty whereas shorter duration was chosen to indicate certainty. In addition, the effect of the locative pronoun was found on duration. The locative "burada" (here) had the shortest duration. High pitch range was chosen to indicate both certainty and uncertainty. Lower pitch range was chosen to indicate neutrality. The locative "şurada" combined with the adverb "kesinlikle" had highest pitch, duration, and pitch range.

For the utterances (i.e., locative pronouns, and epistemic adverbs), high pitch was used to express certainty wheras lower pitch was used to express neutral modality. Lowest pitch was used to express uncertain modality (non-significant).

For the utterances (i.e., locative pronouns, and epistemic adverbs), shorter durations were used when expressing certainty. It was longer for neutral and longest for uncertain modality (non-significant).

For the utterances (i.e., locative pronouns), high pitch range was used to express certainty (non-significant).

For the utterances (i.e., epistemic adverbs), high pitch range was used to express the modality towards neutral modality (non-significant).

When locative pronouns were modified with epistemic adverbs, the effect of locative pronouns disappeared and duration did not differ significantly.

CHAPTER 4

EXPERIMENT 2

4.1 Aim

Listeners infer confidence level of the speakers by using different cues. There are definite nonverbal cues that are used by speakers in production of certainty and uncertainty, and are attended to by listeners to perceive speaker confidence. This shows significance of nonverbal cues in perception of both certainty and uncertainty. However, it is not well known that which of these cues features in perception of speaker confidence.

The primary objective of this study is to investigate and demonstrate the effect of cognitive state such as epistemic stance of the speaker, which is conveyed intonationally and lexically on the speaker confidence perception of the listeners. Are some acoustic properties in speech more reliable to distinguish between certainty and uncertainty of the speaker? Specifically, we want to determine how differences among prosodic elements such as mean pitch, pitch range, and duration of an utterance affect perception of epistemic confidence (i.e., certain and uncertain) of speakers. We want to know acoustic properties and their reference to perception of epistemic confidence of speakers.

We wanted to explain the link between prosody and pragmatic meaning. The study was conducted in order to detect how speakers regulate the acoustic properties (i.e., mean pitch, pitch range, and duration) during speech and whether this regulation will be affected by lexical items such as locative pronouns (burada, şurada, orada) and epistemic adverbs (kesinlikle, galiba) under certain and uncertain conditions. Does spatial semantic provide a cue about epistemic confidence of the speaker in a dialogue? Do adverbs provide a cue about epistemic confidence of the speaker in a dialogue?

Since there is lack of definition of locative pronouns in Turkish, the study aimed to define semantics of locative pronouns. We linked speaker confidence via linguistic encoding to proximal and distal locative pronouns. In languages such as Jahai and Dutch cognitive accessibility is assocaited with reachability and retrieval of mental content to mind, respectively (Burenhult, 2003; Piwek, Beun, & Cremers, 2007). We wanted to observe the relationship between access to knowledge and physical distance. Since it is known that semantic spatial encoding of “bu” takes into account the distance of the referent thus used for proximality (Küntay & Özyürek, 2006).

The stimuli in the perception experiment were composed of the corpus of utterances elicited in the production study. The participants listened to short audio clips which were collected in Experiment 1 (see Table 4.1, Table 4.2, & Table 4.3), and evaluated them in terms of how certain or uncertain the producers of the utterances were of the information they conveyed. The aim of this experiment was thus to determine which characteristics are indicative of certainty and uncertainty based on the perception of the listeners.

4.2 Participants

The present experiment included thirty-two (16 females and 16 males) participants who were native speakers of Turkish (*M* = 27,187, *SD* = 3,505). In order to find voluntary participants, the snowball sampling method was employed. All participants were Master and PhD students of the Informatics Institute at Middle East Technical University (METU), Ankara.

4.3 Materials & Design

The sessions were held in the Cogs Lab at the Informatics Institute of METU. The experiment was run on an Intel® Core™ i7 Lenovo-PC running Windows 8.1 64-bit with 15.6" LCD screen with a resolution of 1366 X 768 pixels and a refresh rate of 60p Hz. Open-source software OpenSesame version 3.1. (Mathôt et al., 2012) was used to create and present the experiment. The participants listened to auditory stimuli via earphones. The stimuli in the experiment consisted of audio clips collected from the 7 female native speakers of Turkish in the first experiment of the study. These audio clips were recorded when the speakers replied to a question both under certainty and uncertainty prosodic conditions by using adverbs and locative pronouns. Because of the within-subjects design, all of the participants were subjected to all four blocks of the experiment which were administered in the same order. However, items of the blocks were presented in random order.

In the first block of the experiment, there were fourty-two auditory stimuli. Locative pronouns, namely "burada" (here), "orada" (there), and "şurada" (there) were expressed both in certain and uncertain prosody by the previously mentioned 7 female speakers. The locative pronouns were presented in the first block. The evaluation question was "What kind of response was the audio you listened to?"(Dinlediğiniz ses nasıl bir cevaptı?). In the two alternative forced-choice method, the two options were a) certain (emin) or b) uncertain (emin olmayan) (see Figure 4.5).

In the second block of the experiment, there were fourteen auditory stimuli. The adverb "kesinlikle" (certainly) was expressed both in neutral and certain prosody by the 7 female speakers in the second block of the experiment. The evaluation question was the same. However, in the two alternative forced-choice method the two options differed and were a) neutral (nötr) or b) certain (emin) (see Figure 4.9).

The adverb “galiba” (probably) was expressed both in neutral and uncertain prosody by the 7 female speakers in the third block of the experiment. The same evaluation question was asked with two alternative options a) neutral (nötr) or b) uncertain (emin olmayan) in the two alternative forced-choice method (see Figure 4.12).

The fourth block of the experiment consisted of two parts. The stimuli in the first part were comprised of twenty-one audio clips from the 7 female speakers. Each of the speakers had 3 audio clips in which locative pronouns were modified by an adverb. The combinations were “kesinlikle burada” (It is certainly here), “kesinlikle orada” (It is certainly there), and “kesinlikle şurada” (It is certainly there). A 7-point Likert scale was used to evaluate them. All the numbers were available in the scale yet only the numbers 4 (neutral) to 7 (very certain) were labeled (see Figure 4.14). In the second part of the fourth block the same locative pronouns were modified by another adverb. The stimuli were “galiba burada” (It is probably here), “galiba orada” (It is probably there), and “galiba şurada” (It is probably there). Although the same scale was used in the second part of the fourth block, only the numbers 1 (not certain at all) to 4 (neutral) were labeled (see Figure 4.16). Even though the participants received the four blocks in the same order, all the auditory stimuli in the blocks were presented in random order.

The study was approved by METU ethics committee (see Appendix A).The participants were invited into the Cogs Lab and seated at a table. The Lenovo-PC was placed on the table and earphones were plugged in. The experimenter ran OpenSesame (Mathôt et al., 2012) to present the stimuli and explained to the participants that they would see and read specific information before each block on the laptop screen. The program required a Subject ID, which was typed by the experimenter. The participant number, sex, and age were concatenated to attain a Subject ID (e.g., 14M23). A welcome message was displayed when the experiment started. The participants were requested to confirm the consent form, in which the purpose and the ethical concerns of the experiment were explained (see Appendix E). To be able to start the experiment, the participants had to confirm the form by selecting the “continue” button and wear earphones (see Figure 4.1).

Figure 4.1: Informed consent form

An instruction text indicated that the audio clips they would be listening to were made up of the answers given by using adverbs and locative pronouns under certainty and uncertainty conditions. An imaginary scenario, which was used when the answers were recorded, was introduced to the participants. According to the imaginary scenario, a key was hidden. A person, the experimenter, entered the room and asked the question "Where is the key?". Under certainty the speakers knew the place of the key and were sure of it, whereas under uncertainty they did not know the place of the key and were unsure of it. Morever, the participants were informed that the experiment involved four blocks. They were advised to read the specific information displayed in the screen before each block carefully. A keypress was required to continue to the next screen (see Figure 4.2).

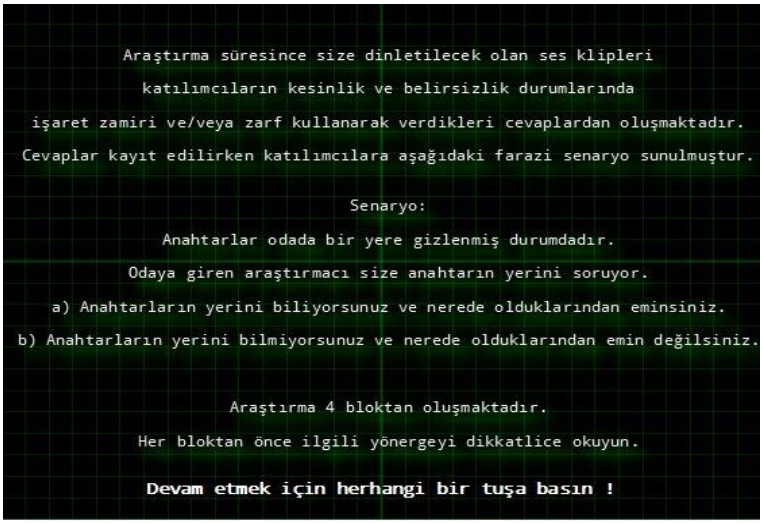

Figure 4.2: The instruction text

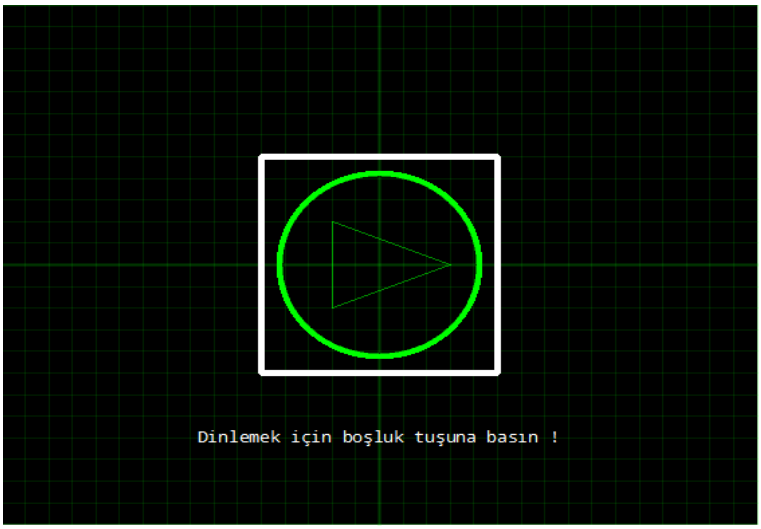

Figure 4.3: Audio play screen in all blocks of the experiment

After all the auditory stimuli had been listened to and rated, a message on the feedback screen informed the participants that this was the end of the experiment. They were thanked for their participation in the experiment. Finally, they terminated the experiment by a keypress (see Figure 4.4). As the last step of the procedure they received the debriefing information (see Appendix F). When the experiment was terminated OpenSesame created a single comma-separated file for each participant and named those files with their Subject ID. The logger recorded all variables as well as the responses. The files were converted into tab-delimited files by the experimenter in order to use them in the data analysis part.

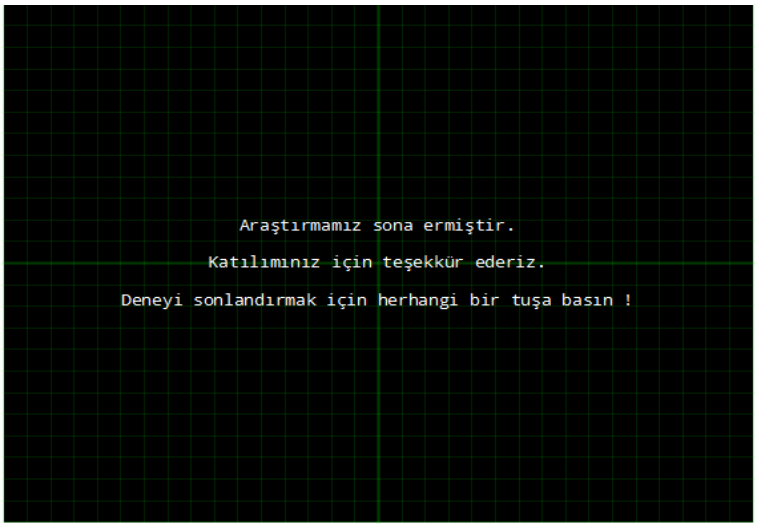

Figure 4.4: The end of the experiment screen

4.5 Data coding

Data coding of responses elicited from the perception experiment was conducted in the following ways.

Speaker confidence (perceived modality) had binary response. For Block I, the participants selected one of the options whether the speaker was certain or uncertain in the perception experiment. The value of 1 was given to the response "certain" while the value of 0 was given to the response "uncertain". For Block II, the participants selected one of the options whether the speaker was certain or neutral in the perception experiment. The value of 1 was given to the response "certain" while the value of 0 was given to the response "neutral". For Block III, the participants selected one of the options whether the speaker was uncertain or neutral in the perception experiment. The value of 1 was given to the response "uncertain" while the value of 0 was given to the response "neutral".

Correctness (accuracy) had binary response. For the blocks I, II, and III, if modality (i.e., certain or uncertain) and perceived modality (i.e., certain or uncertain) were matched, the score of 1, indicating modality was able to be conveyed, was given to the item. If not, the score of 0, indicating modality was not able to be conveyed, was given to the item.

Speaker confidence score (perceived modality score) is the evaluation of items on a 7-point Likert scale. Speaker confidence score had continuous response. For the blocks IV.I and IV.II, the scores were normalized according to the value of 4. Thus, the maximum certainty and uncertainty score a speaker could get was the value of 3 and the minimum certainty and uncertainty score a speaker could get was the value of 0.

The mean pitch of items was calculated and obtained mean pitch value was subtracted from the pitch value of each item. Thus, centered mean pitch was obtained. The mean pitch range of items was calculated and obtained mean pitch range value was subtracted from the pitch range value of each item. Thus, centered pitch range was obtained. The mean duration of items was calculated and obtained mean duration was subtracted from the duration of each item. Thus, centered duration was obtained.

4.6 Analysis

For the blocks I, II, and III, speaker confidence (perceived modality) and correctness (accuracy), the binary responses, were analysed performing a Generalized Linear Mixed Model (GLMM) using glmer ( ) function in lme4 package (Bates, Maechler, Bolker, & Walker, 2015) in the R statistical programming environment for statistical computing and graphics version 3.3.2 (R Core Team, 2017). The terms "generalized" is used for nonnormal distributions and "mixed" is used for random effects in GLMM. Based on the recorded responses that composed categoric variables, GLMM was performed to be able to analyse binary data. GLMM uses logistic link function that is L=log(p/(1-p)). L refers

to an expected response within the function. “Y is 1 with probability p and 0 with probability 1-p” (Dickey, 2010, p. 1).

For the blocks IV.I and IV.II, the responses were analysed performing a Linear Mixed Model (LMM) using lmer ( ) function in lme4 package (Bates, Maechler, Bolker, & Walker, 2015) in the R statistical programming environment for statistical computing and graphics version 3.3.2 (R Core Team, 2017). The term “linear” is used for fixed effects and “mixed” is used for random effects in LMM. Based on the recorded responses that composed a continuous variable, LMM was performed.

When the models were created, bottom-up approach was followed. The factors were added to the intercept model step by step. Constructed models were compared in terms of likelihood. The anova ( ) function was used and the value $\alpha_{LRT} = .05$ was choosen as significance level for likelihood ratio test to be able to select among two models (Matuschek, Kliegl, Vasishth, Baayen, & Bates, 2017). “REML=FALSE” suggested to be added to the model when models are constructed (Pinheiro & Bates, 2000; Bolker et al., 2009). The models that had convergence warnings did not included in the steps any further since this warning is considered to be false positive (Bates et al., 2013). Furthermore, normality assumption was not controlled due to linear mixed model (Gelman & Hill, 2006). When reporting the statistical analysis, the tutorial by Winter (2013) was followed (see Appendix G).

4.7 Block I

The text “Block I” was displayed for 3000 ms by a keypress and the screen switched to the instruction text belonging to the first block. The specific information indicated that locative pronouns which were “burada” (here), “orada” (here), and “şurada” (here) would be heard by the participants. They were required to evaluate the audio clips that they listened to by considering the confidence of the speakers (whether they were sure or unsure where the key was) with the use of the two alternative forced choice method. They pressed the spacebar in order to start with the first block. They listened to each audio clip by pressing the spacebar (see Figure 4.3) and on the following screen they could state their evaluation by selecting one of the two options - whether their response was certain (i.e., the speaker was sure) or uncertain (i.e., the speaker was unsure where the key was) (see Figure 4.5). At the end of the block, the participants received a message informing them about the ending. They continued with the second block by a keypress.

Figure 4.5: The evaluation question with the two alternative forced-choice method in Block I

4.7.1 Data preparation & Analysis

An Excel file was created for Block 1 with the columns participant number, item number, speaker number, locative, modality, duration, mean pitch, pitch range, centered duration, centered mean pitch, centered pitch range, and speaker confidence (perceived modality) and correctness (accuracy).

All of the utterances elicited from the Experiment 1 (production study) in which for each of the 7 speakers there were six utterances (2 modalities * 3 locatives), for a total of 42 utterances were involved in the first block of the Experiment 2 (perception study). A total of 1344 responses were recorded from the perception study of Block I (42 responses x 32 participants). The recorded responses composed a categoric variable that is speaker confidence (perceived modality) with two levels (i.e., certain or uncertain). Moreover, the dependent variable that is correctness (accuracy) was created based on modality and speaker confidence (perceived modality) as a categorical variable with two levels (i.e., correct or incorrect). Therefore, GLMM was performed to be able to analyse binary data. The fixed factors (independent variables) were acoustic properties (i.e., mean pitch, duration, and pitch range), locative (with three levels: burada (here), şurada (there), and orada (there), and modality (with two levels: certain, and uncertain). The random factors were participant.no and item.

4.7.2 Result

Speaker Confidence (Perceived Modality)

At the end of the fixed and random structuring the selected models are as the following:

Formula 1: Speaker_Confidence ~ C.Dur + CPRange + (1 | Participant.no) + (1 | item)

Standard deviation is a measure of the variability for the random effects participants and items within the model. Items "item" has more variability (*SD* = 1.0617) than participants "participant.no" (*SD* = 0.5214).

Table 4.1: The coefficient table for the effect of duration and pitch range

Fixed effects:				
	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	0.747832	0.202252	3.698	0.000218 ***
C.Dur	-0.009937	0.002020	-4.920	8.65e-07 ***
CPRange	0.008709	0.003292	2.646	0.008150 **

The estimate value for intercept was transformed using inverse-logit function. The function invlogit ( ) yielded the value of 0.6787061, indicating that when centered duration is zero (the value is mean duration) and centered pitch range is zero (the value is mean pitch range), the probability of the estimation of the speaker confidence (perceived modality) as certain is 0.67 (b = 0.747832, SE = 0.202252, z = 3.698, p < .001). The coefficient "C.Dur" (centered duration) is the slope for the continuous effect of duration. Minus 0.009937 means that when centered duration increases one unit, the probability of the estimation of the speaker confidence (perceived modality) as certain decreases. Statistically significant result indicates that certainty is less for longer durations than it is for shorter durations (b = -0.009937, SE = 0.002020, z = -4.920, p < .001). The coefficient "CPRange" (centered pitch range) is the slope for the continuous effect of duration. 0.00870 means that when pitch range increases one unit, the probability of the estimation of the speaker confidence (perceived modality) as certain increases. Statistically significant result indicates that certainty is more for higher pitch range (b = 0.008709, SE = 0.003292, z = 2.646, p < .01). R^2 value 0.3110569, indicating that about 31.1% of variance is able to be explained by the selected model.

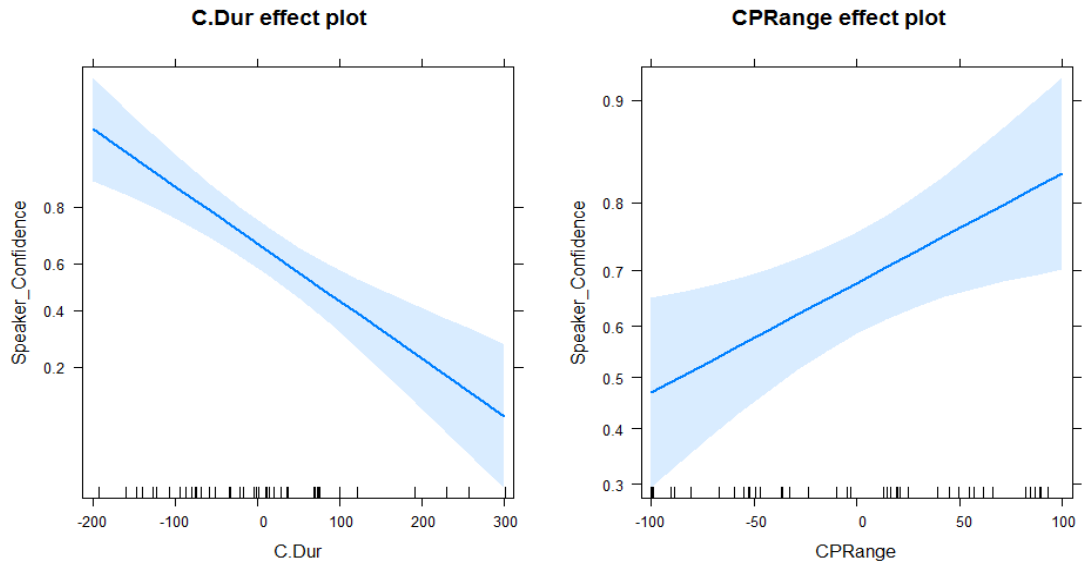

Figure 4.6: Showing the effect of duration and pitch range on speaker confidence (perceived modality)

Formula 2: Speaker_Confidence ~ C.Dur + Locative + (1 | Participant.no) + (1 | item)

Standard deviation is a measure of the variability for the random effects participants and items within the model. Items "item" has more variability (*SD* = 1.0589) than participants "participant.no" (*SD* = 0.5215).

Table 4.2: The coefficient table for the effect of duration and locative pronouns

Fixed effects:				
	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	0.492637	0.344574	1.430	0.153
C.Dur	-0.009829	0.002043	-4.811	1.5e-06 ***
Locative2	1.058376	0.535340	-5.751	0.048 *
Locative3	-0.286503	0.438634	-0.653	0.514

The estimate value for intercept was transformed using inverse-logit function. The function invlogit ( ) yielded the value of 0.6207274, indicating that when centered duration is zero (the value is mean duration) and the locative is "Locative1" (şurada

(there)), the probability of the estimation of the speaker confidence (perceived modality) as certain is 0.62 (b = 0.49267, SE = 0.344574, z = 1.430, p = 0.153). The coefficient "C.Dur" (centered duration) is the slope for the continuous effect of duration. Minus 0.009829 means that when centered duration is increased one unit, the probability of the estimation of the speaker confidence (perceived modality) as certain is decreased by 0.009829. Statistically significant result indicates that certainty is less for longer durations than it is for shorter durations (b = -0.009829, SE = 0.002043, z = -4.811, p < .001). The coefficients "Locative2" (burada (here)) and "Locative3" (orada (there)) are the slopes for the categorical effect of locative. 1.058376 means that when going from "Locative1" (şurada (there)) to "Locative2" (burada (here)), the probability of the estimation of the speaker confidence (perceived modality) as certain increases by 1.058376. Statistically significant result indicates that certainty is less for "Locative1" (şurada (there)) than it is for "Locative2" (burada (here)) (b = 1.058376, SE = 0.535340, z = -5.751, p < .001). Minus 0.286503 means that when going from "Locative1" (şurada (there)) to "Locative3" (orada (there)), the probability of the estimation of the speaker confidence (perceived modality) as certain decreases by 0.286503. Although the result is not statistically significant, indicating that certainty is more for "Locative1" (şurada (there)) than it is for "Locative3" (orada (there)) (b = -0.286503, SE = 0.438634, z = -0.653, p = 0.514). R^2 value 0.3109982, indicating that about 31% of variance is able to be explained by the selected model.

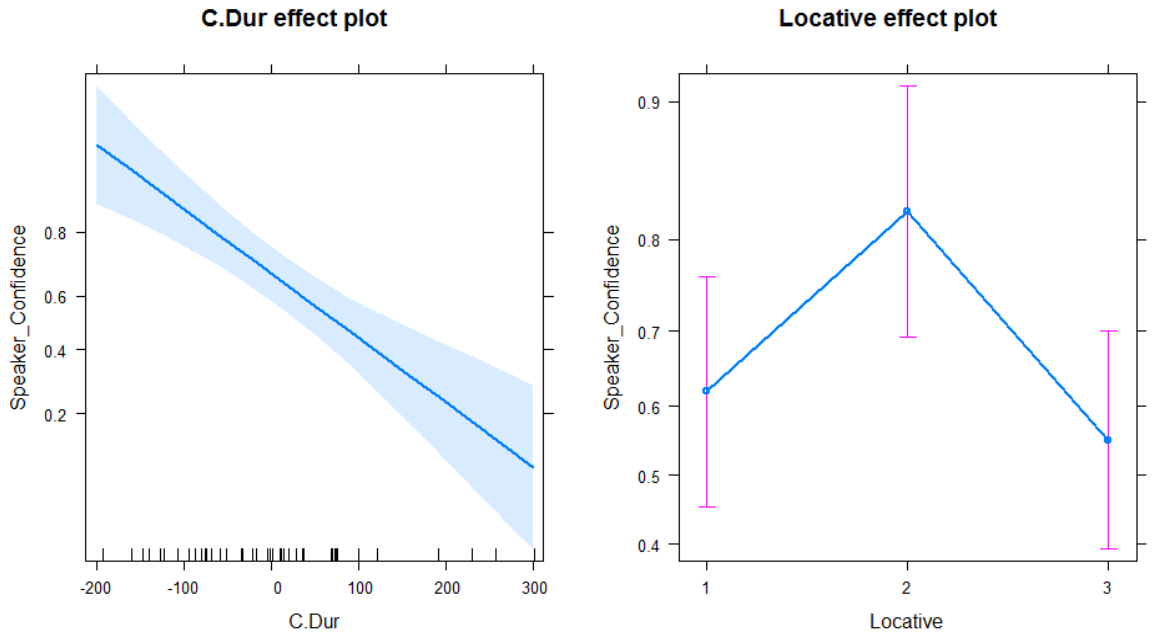

Figure 4.7: Showing the effect of duration and locative pronouns on speaker confidence (perceived modality)

Correctness (Accuracy)

At the end of the fixed and random structuring the selected model is as the following:

Formula: Correctness ~ CMPitch + Modality + (1 | Participant.no) + (1 | item)

Standard deviation is a measure of the variability for random effects; participants and items within the model. Items "item" has more variability (*SD* = 0.8147) than participants "participant.no" (*SD* = 0.3121).

Table 4.3: The coefficient table for the effect of mean pitch and modality

Fixed effects:				
	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	1.753121	0.226378	7.744	9.62e-15 ***
CMPitch	-0.02565	0.006103	-4.203	2.63e-05 ***
Modality2	-1.73135	0.301066	-5.751	8.89e-09 ***

The estimate value for intercept was transformed using inverse-logit function. The function invlogit ( ) yielded the value of 0.8523302, indicating that when centered mean pitch is zero (the value is mean pitch) and the modality is certain, the probability of the correct estimation of the modality of an item (whether it is certain or not) is 0.85 (b = 1.753121, SE = 0.226378, z = 7.744, $p < .001$). The coefficient "CMPitch" (centered mean pitch) is the slope for the continuous effect of pitch value. Minus 0.02565 means that when centered mean pitch is increased one unit, the probability of the correct estimation of the modality of an item (whether it is certain or not) decreases. Statistically significant result indicates that correctness is less in high mean pitch than in low mean pitch (b = -0.02565, SE = 0.006103, z = -4.203, $p< .001$). The coefficient "Modality2" (uncertain) is the slope for the categorical effect of modality. Minus 1.731348 means that when going from "certain" to "uncertain", the probability of the correct estimation of the modality of an item (whether it is certain or not) decreases. Statistically significant result indicates that correctness is less in uncertain modality than in certain modality (b = -1.73135, SE = 0.301066, z = -5.751, $p < .001$). R^2 value 0.2566408, indicating that about 25.6% of variance is able to be explained by the selected model.

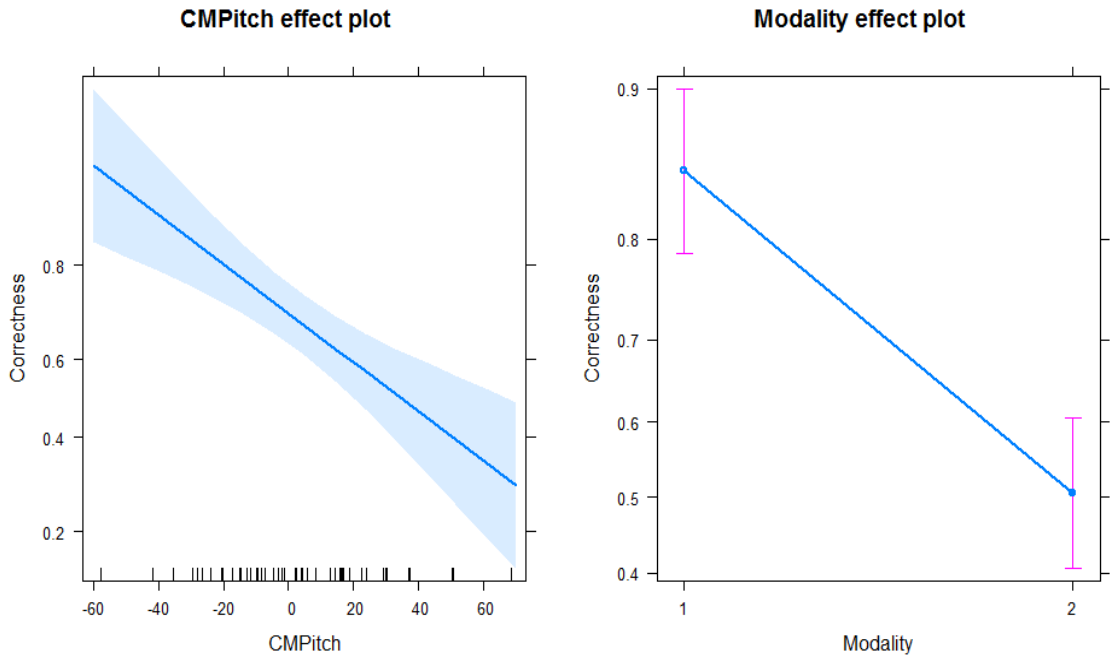

Figure 4.8: Showing the effect of mean pitch and modality on correctness (accuracy)

4.7.3 Discussion

According to perception of the speaker confidence the locative pronouns with shorter durations determined perception of the confidence of speaker as certain whereas longer durations determined it as uncertain. There are studies providing our results. Longer duration, and slower speech rate were perceived as unconfident whereas faster speech rate were perceived as confident (Jiang & Pell, 2017). Moreover, Dral, Heylen, & op den Akker (2011) suggested that low speed of talking as property of uncertainty. Locative pronouns with wider pitch range determined perception of the confidence of speaker as certain whereas narrower ranges determined it as uncertain. In a previous study, the dynamic change of pitch helped to distinguish between confidence and unconfidence. With respect to fundamental frequency range, the utterances which were evaluated as confident had highest value (Jiang & Pell, 2017). In addition, another result indicated that locative “burada” (here) mostly perceived as certain by the participants. It is followed by “şurada” (here) and “orada” (there). Construal level theory claims that people construe events as well as objects differently in terms of their spatio-temporal distance from themselves (Trope & Liberman, 2010). It could thus be that distance is also sensitive to "certainty" with proximal objects (burada) being construed as more certain than distal objects (şurada, orada). Likewise, cognitive accessibility (Burenhult, 2003; Kahneman, 2003) might be deteminer on the perception of distal and proximal locative pronouns. However, we approach this result with caution since this effect may be due to lenght of stimuli itself. Duration of the stimuli effects how pitch contour is perceived (Chen, Zhu & Wayland, 2017).

With respect to accuracy, independently of modality of the items, lower pitch value facilitated correct estimation of the speaker confidence. As the mean pitch value increased, accuracy lowered. The expectation of listeners and expression of speakers were differed with respect to pitch. The number of correct match is greater for certain modality. The way speakers voiced certain modality were in accordance with perception of listeners.

4.8 Block II

The text “Block II” was displayed for 3000 ms and the screen switched to the instruction text for the second block. The specific information indicated that the adverb “kesinlikle” (certainly) would be heard. The participants were required to evaluate the audio clips they listened to by considering the confidence of the speakers (neutral or sure) with the use of the two alternative forced choice method. The block was initiated by pressing the spacebar. Each audio clip was played by pressing the spacebar (see Figure 4.3) and after the audio was played the evaluation question was displayed with a two aternative choice - whether the audio played was neutral (i.e., the speaker was neutral) or certain (i.e., the speaker was sure where the key was) (see Figure 4.9). When all audio clips were played, the feedback screen was displayed and participants were informed about the ending of the second block.

Figure 4.9: The evaluation question with the two alternative forced-choice method in Block II

4.8.1 Data preparation & Analysis

An Excel file was created for Block 2 with the columns as it was prepared for Block 1 with an additional column “adverb”. Speaker confidence (perceived modality) had binary response (i.e., certain or neutral).

All of the utterances elicited from the Experiment 1 (production study) in which for each of the 7 speakers there were two utterances (2 modalities * 1 adverb), for a total of 14 utterances were involved in the second block of the Experiment 2 (perception study). A total of 448 responses were recorded from the perception study of Block II (14 responses x 32 participants). The recorded responses composed a categoric variable that is speaker confidence (perceived modality) with two levels (i.e., certain or neutral). Therefore, GLMM was performed to be able to analyse binary data. The fixed factors (independent variables) were acoustic properties (i.e., mean pitch, duration, and pitch range), and modality (with two levels: certain and neutral). The random factors were participant.no and item.

4.8.2 Result

Speaker Confidence (Perceived Modality)

At the end of the fixed and random structuring the selected model is as the following:

Formula: Speaker_Confidence ~ Modality + (1 | Participant.no) + (1 | item)

Standard deviation is a measure of the variability for the random effects participants and items within the model. Items "item" has more variability (*SD* = 0.8824) than participants "participant.no" (*SD* = 0.2583).

Table 4.4: The coefficient table for the effect of modality

Fixed effects:				
	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	1.3545	0.3830	3.536	0.000406 ***
Modality3	-1.9079	0.5291	-3.606	0.000311 ***

The estimate value for intercept was transformed using inverse-logit function. The function invlogit ( ) yielded the value of 0.7948643, indicating that when the modality is certain the probability of the estimation of the speaker confidence (perceived modality) as certain is 0.79 (b = 1.3545, SE = 0.3830, z = 3.536, p < .001). The coefficient "Modality3" (neutral) is the slope for the categorical effect of modality. Minus 1.9079 means that when going from "Modality1" (certain) to "Modality3" (neutral), the probability of the estimation of the speaker confidence (perceived modality) as certain decreases. Statistically significant result indicates that certainty is less for "Modality3" (neutral) than it is for "Modality1" (certain) (b = -1.9079, SE = 0.5291, z = -3.606, p <

.001). R^2 value 0.3092213, indicating that about 30.9% of variance is able to be explained by the selected model.

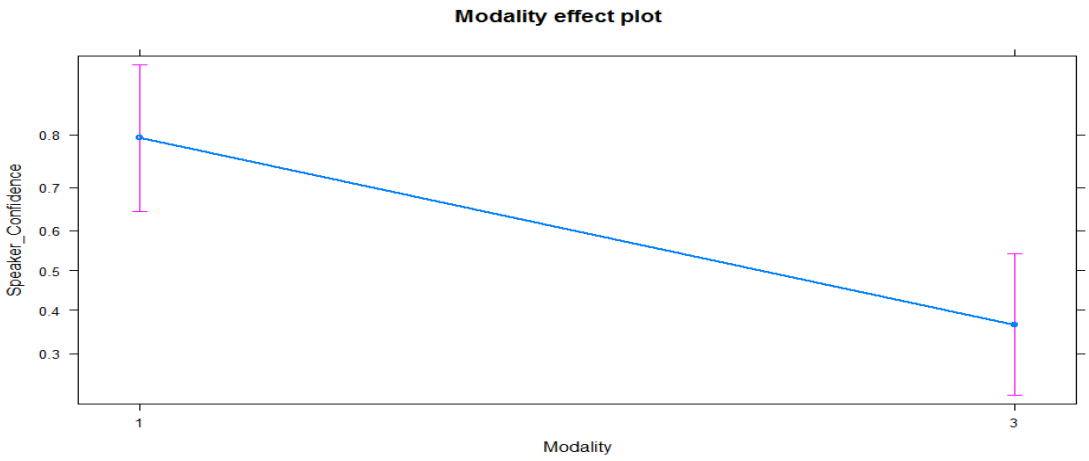

Figure 4.10: Showing the effect of modality on speaker confidence (perceived modality)

Correctness (Accuracy)

At the end of the fixed and random structuring the selected model is as the following:

Formula: Correctness ~ CMPitch + (1 | Participant.no) + (1 | item)

Standard deviation is a measure of the variability for the random effects; participants and items within the model. Items "item" has more variability (*SD* = 0.7778) than participants "participant.no" (*SD* = 0.2798).

Table 4.5: The coefficient table for the effect of mean pitch

Fixed effects:				
	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	0.94334	0.24444	3.859	0.000114 ***
CMPitch	0.02348	0.01052	2.231	0.025659 *

The estimate value for intercept was transformed using inverse-logit function. The function invlogit ( ) yielded the value of 0.7197738, indicating that when centered mean pitch is zero (the value is mean pitch), the probability of the correct estimation of the modality of an item (whether it is certain or neutral) is 0.71 (b = 0.94334, SE = 0.24444, z = 3.859, p < .001). The coefficient "CMPitch" (centered mean pitch) is the slope for

the continuous effect of pitch value. 0.02348 means that when centered mean pitch increases one unit, the probability of the correct estimation of the modality of an item (whether it is certain or neutral) increases. Statistically significant result indicates that correctness is more in high mean pitch than in low mean pitch (b = 0.023489, SE = 0.01052, z = 2.231, p < .05). R^2 value 0.2118217, indicating that about 21.1% of variance is able to be explained by the selected model.

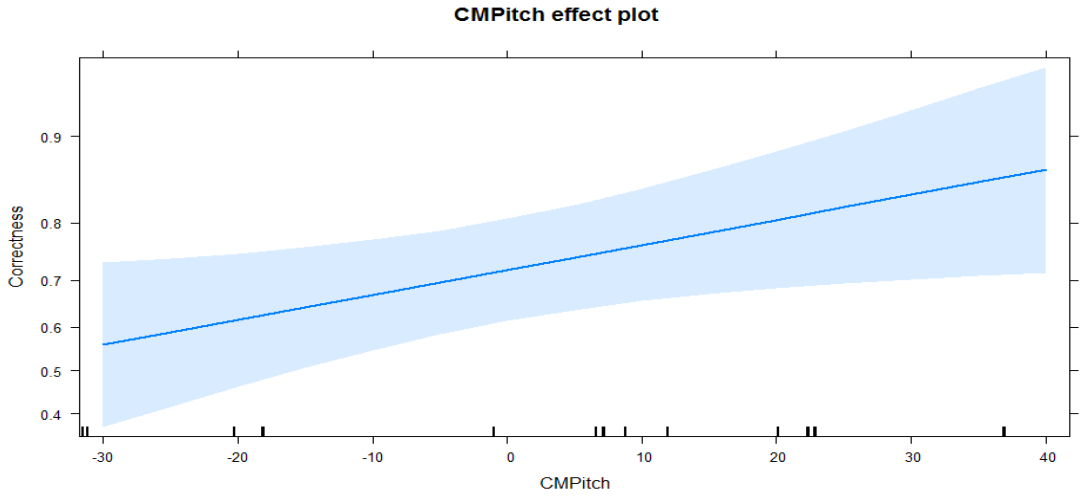

Figure 4.11: Showing the effect of mean pitch on correctness (accuracy)

4.8.3 Discussion

According to perception of the speaker confidence (perceived modality) none of the parameters (mean pitch, duration, and pitch range) determined perception of the confidence of speaker as certain whereas modality determined it as uncertain. Participants already knew that it is used for certainty. The semantic of the word was found to be effective. Labeled modality determined the perception. However, there might be another paralinguistic content that empahisized modality of the item. In this case lexical content surpassed the acoustic properties as it was observed in the studies with children (see Friend, 2000; Friend & Bryant, 2000; Morton & Trehub, 2001; Solomon & Ali, 1972, as cited in Friend, 2003) and this adultlike behavior occured in studies in which paralanguage and language contradicted.

With respect to accuracy, independently of modality of the items, higher pitch value facilitated correct estimation of the speaker confidence. As the mean pitch value increased, accuracy rised. The expectation of listeners and expression of speakers were more likely to be in congruity with respect to pitch. The number of correct match is greater for certain modality. The way speakers voiced certain modality were in accordance with perception of listeners. We did not observe a specific pattern of pitch within modalities based on the evaluation of the listeners.

4.9 Block III

The text “Block III” was displayed and stayed for 3000ms but changed when the participants pressed a key to continue. The switched screen displayed the specific information indicating that the adverb “galiba” (probably) would be heard. The participants were required to evaluate the audio clips they listened to by considering the confidence of the speakers (neutral or unsure) with the use of the two alternative forced choice method - whether the audio played was neutral (i.e., the speaker was neutral) or uncertain (i.e., the speaker was unsure where the key was). They continued with the third block by pressing on the spacebar. They pressed the spacebar to listen to the audio clips as well (see Figure 4.3). Two alternative forced choice question was displayed with the options whether the audio played was neutral (i.e., the speaker was neutral) or unsure (i.e., the speaker was unsure where the key was) (see Figure 4.12). The feedback screen notified participants about the ending of the third block and the requirement of a keypress to continue with the next block.

Figure 4.12: The evaluation question with the two alternative forced-choice method in Block III

4.9.1 Data preparation & Analysis

An Excel file was created for Block 3 with the columns as it was prepared for Block 2. Speaker confidence had binary response (i.e., uncertain and neutral).

All of the utterances elicited from the Experiment 1 (production study) in which for each of the 7 speakers there were two utterances (2 modalities * 1 adverb), for a total of 14 utterances were involved in the third block of the Experiment 2 (perception study). A total of 448 responses were recorded from the perception study of Block III (14 responses x 32 participants). The recorded responses composed a categoric variable that is speaker confidence with two levels (i.e., uncertain or neutral). Moreover, the

dependent variable that is correctness was created based on modality and speaker confidence as a categorical variable with two levels (i.e., correct or incorrect). Therefore, GLMM was performed to be able to analyse binary data. The fixed factors (independent variables) were acoustic properties (i.e., mean pitch, duration, and pitch range), and modality (with two levels: uncertain and neutral). The random factors were participant.no and item.

4.9.2 Result

Speaker Confidence (Perceived Modality)

At the end of the fixed and random structuring the selected models are as the following:

Formula: Speaker_Confidence ~ CMPitch + C.Dur + (1 | Participant.no) + (1 | item)

Standard deviation is a measure of the variability for the random effects participants and items within the model. Items "item" has less variability (*SD* = 0.6103) than participants "participant.no" (*SD* = 0.6299).

Table 4.6: The coefficient table for the effect of mean pitch and duration

Fixed effects:				
	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	0.47431	0.229231	0.0207	0.83608
CMPitch	0.018720	0.007336	2.552	0.01071 ***
C.Dur	0.005825	0.002146	2.714	0.00665 **

The estimate value for intercept was transformed using inverse-logit function. The function invlogit ( ) yielded the value of 0.6164034, indicating that when centered mean pitch is zero (the value is mean pitch) and duration is zero (the value is mean duration), the probability of the correct estimation of the speaker confidence as neutral is 0.61 (b = 0.47431, SE = 0.229231, z = 0.83608, p = 0.83608). The coefficient "CMPitch" (centered mean pitch) is the slope for the continuous effect of pitch value. 0.018720 means that when centered mean pitch increases one unit, the probability of the estimation of the speaker confidence (perceived modality) as neutral increases. Statistically significant result indicates that neutrality is more in high mean pitch than in low mean pitch (b = 0.018720, SE = 0.007336, z = 2.552, p < .05). The coefficient "C.Dur" (centered duration) is the slope for the continuous effect of duration. 0.005825 means that when centered duration increases one unit, the probability of the estimation of the speaker confidence (perceived modality) as neutral increases. Statistically significant result indicates that neutrality is more in longer durations than in shorter

duration (b = 0.005825, SE = 0.002146, z = 2.714, p < .01). R^2 value 0.3073622, indicating that about 30.7% of variance is able to be explained by the selected model.

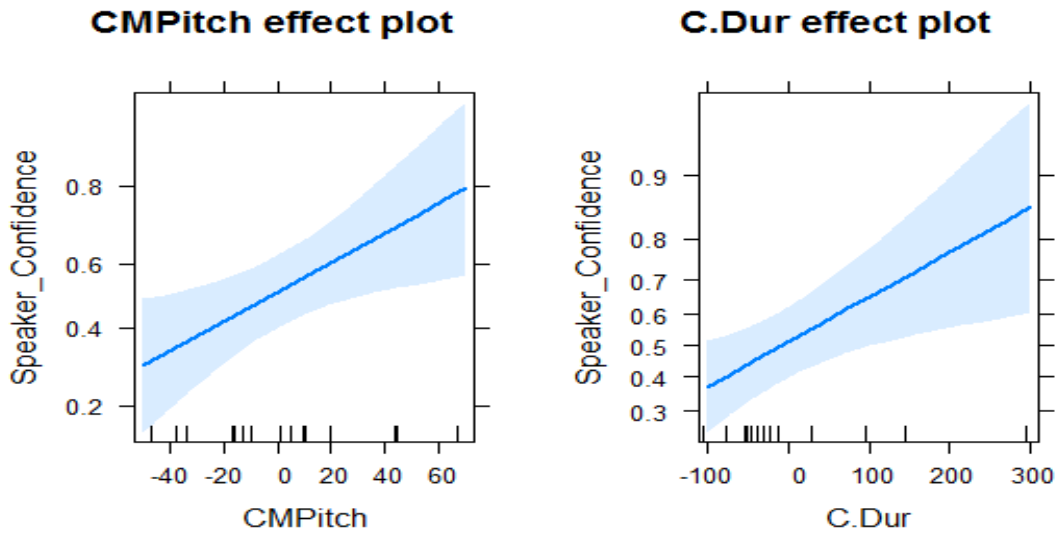

Figure 4.13: Showing the effect of mean pitch and duration on speaker confidence (perceived modality)

Correctness (Accuracy)

At the end of the fixed structuring the selected model is as the following:

Correctness ~ 1 + (1 | Participant.no) + (1 | item)

Standard deviation is a measure of the variability for random effects; participants and items within the model. Items "item" has more variability (SD = 0.9358) than participants "participant.no" (SD = 0.0000).

Table 4.7: The coefficient table for the intercept model

Fixed effects:				
	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	0.4528	0.2723	1.663	0.0964

The estimate value for intercept was transformed using inverse-logit function. The function invlogit ( ) yielded the value of 0.6113048, when all factors are zero (mean value) the probability of the correct estimation of the modality of an item (whether it is

certain or not) is 0.61 (b = 0.4528, SE = 2723, z = 1.663, p > .05). R^2 value 0.1804197, indicating that about 18% of variance is able to be explained by the selected model.

4.9.3 Discussion

According to perception of the speaker confidence longer durations determined perception of the confidence of speaker as uncertain whereas shorter durations determined it as neutral. There are studies providing our results. Longer duration, and slower speech rate were perceived as unconfident whereas faster speech rate were perceived as confident (Jiang & Pell, 2017). Moreover, Dral, Heylen, & op den Akker (2011) suggested that low speed of talking as property of uncertainty. Moreover, high pitch determined perception of the confidence of speaker as uncertain whereas lower pitch determined it as neutral. This result is contrary to the findings of Ipek & Jun (2013) suggested that English intonation differs from that of Turkish since rising pitch is employed to convey certainty. However, there are also other studies supporting our finding with respect to perception of uncertainty. High and/or rising fundamental frequency conveys politeness and lack of confidence while low and/or falling fundamental frequency conveys aggression and confidence (Bolinger, 1978; Ohala, 1984). Further, rising intonation and lack of confidence are adapted (Smith & Clark, 1993).

With respect to accuracy, we could not observe the effect of neither modality nor acoustic features. There must have been another feature which is not the scope of this study. However, intonation contour should have been examined whether they signaled uncertainty in utterances that generate our corpus.

4.10 Block IV.I

The text “Block IV.I” stayed for the duration of 3000ms on the screen and switched to the next screen when participants pressed spacebar. The information related to first part of the fourth block explained that locative pronouns modified by an adverb “kesinlikle” (certainly) would be presented. These were “kesinlikle burada” (It is certainly here), “kesinlikle orada” (It is certainly there), and “kesinlikle şurada” (It is certainly there). The screen was switched to the auditory stimuli screen by pressing spacebar. The participants were requried to listen each audio clip by pressing spacebar and evaluate them on a 7-point Likert scale between 4 (neutral) and 7 (very certain) by considering how sure the speakers were in their answers. After the participants had rated the audio clip on the Likert scale by clicking on one of the options, they clicked the “next” botton in order to switch the screen and pressed spacebar to listen to the next audio clip (see Figure 4.3 & Figure 4.14). When all audio clips had been played and rated, the message informing them about the ending of the first part of the fourth block was displayed on the feedback screen.

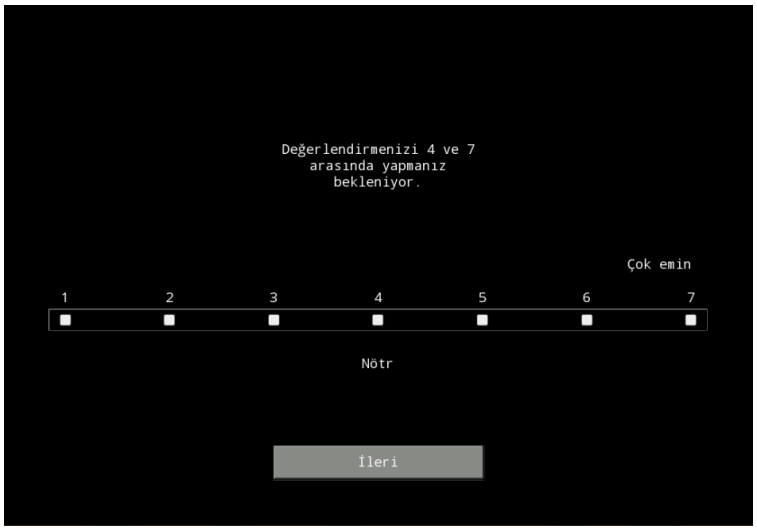

Figure 4.14: The evaluation on a 7-point Likert scale in Block IV.I

4.10.1 Data preparation & Analysis

An Excel file was created for Block 4.1 with the columns as it was prepared for Block 1 with an additional column "adverb + locative" instead of adverb. Speaker confidence score had continuous response. The participants evaluated items on a 7-point Likert scale between 4 (neutral) and 7 (very certain) by considering how sure the speakers were in their answers in the perception experiment. The scores were normalized according to the value of 4. Thus, the maximum certainty score a speaker could get was the value of 3 and the minimum certainty score a speaker could get was the value of 0.

All of the utterances elicited from the Experiment 1 (production study) in which for each of the 7 speakers there were three utterances (1 modality * adverb + 3 locative), for a total of 21 utterances were involved in the first part of the fourth block of the Experiment 2 (perception study). A total of 672 responses were recorded from the perception study of Block IV.I (21 responses x 32 participants). The recorded responses composed a continuous variable that is speaker confidence (certainty score). Therefore, LMM was performed to be able to analyse continous data. The fixed factors (independent variables) were acoustic properties (i.e., mean pitch, duration, and pitch range), adverb + locative (with three levels: "Kesinlikle burada" (It is certainly here), "Kesinlikle şurada" (It is certainly there), and "Kesinlikle orada" (It is certainly there)), and modality (with one level: certain). The random factors were participant.no and item.

4.10.2 Result

At the end of the fixed and random structuring the selected model is as the following:

Formula: Speaker_Confidence_Score ~ C.Dur + CPRange + (1 | Participant.no) + (1 | item)

Standard deviation is a measure of the variability for the random effects participants and items within the model. Participants had more variability "participant.no" (*SD* = 0.4553) than items "item" (*SD* = 0.3751).

Table 4.8: The coefficient table for the effect of duration and pitch range

Fixed effects:				
	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.8131598	0.2921629	2.783	0.01029 *
C.Dur	-0.0009803	0.0002912	-3.367	0.00292 **
CPRange	0.0044163	0.0014889	2.966	0.00742 **

The estimate value for intercept indicates that when centered duration is zero (the value is mean duration) and centered pitch range is zero (the value is mean pitch range), predicted speaker confidence score for certainty is 0.8131598 (b = 0.8131598, SE = 0.2921629, t = 2.783, $p < .05$). The coefficient "C.Dur" (centered duration) is the slope for the continuous effect of duration. Minus 0.0009803 means that when centered duration is increased one unit, speaker confidence score for certainty is decreased by 0.000980. Statistically significant result indicates that speaker confidence score for certainty is higher when duration is shorter *(*b = -0.0009803, SE = 0.0002912, t = -3.367, $p < .01$). The coefficient "CPRange" (centered pitch range) is the slope for the continuous effect of pitch range. 0.0044163 means that when pitch range is increased one unit, speaker confidence score for certainty is increased by 0.0044163. Statistically significant result indicates that speaker confidence score is higher when pitch range is higher (b = 0.0044163, SE = 0.0014889, t = 2.966, $p < .01$). R^2 value 0.4360158, indicating that about 43.6% of variance is able to be explained by the selected model.

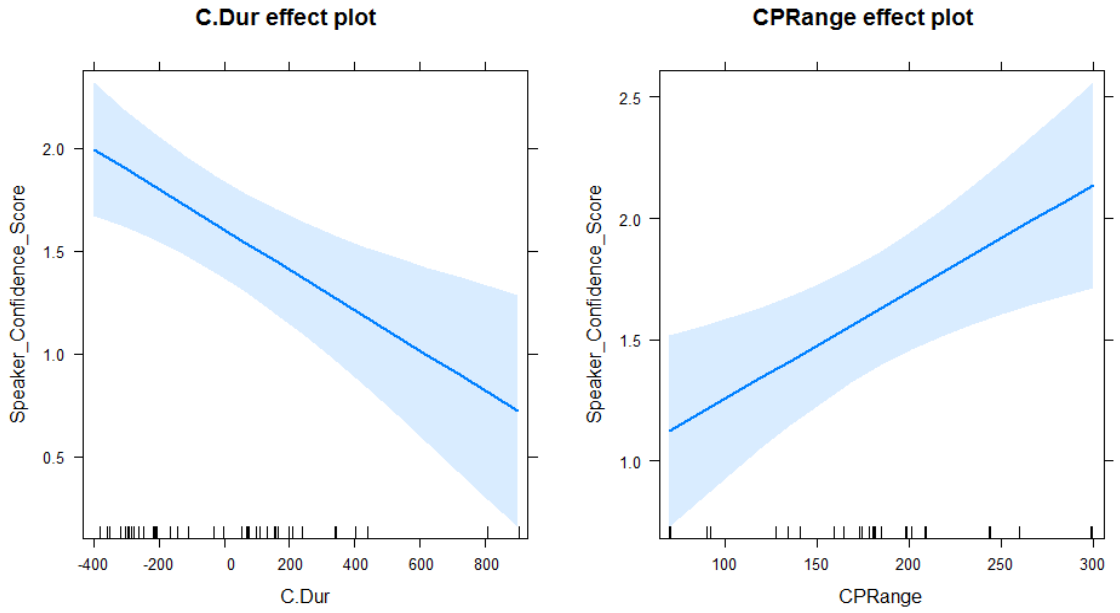

Figure 4.15: Showing the effect of duration and pitch range on speaker confidence score

4.10.3 Discussion

For locative pronouns modified by epistemic adverb "kesinlikle" (certainly), listeners perceived the items more certain when they had shorter total duration. This result is in line with the studies suggests that utterances with longer duration and slower speech rate were perceived as unconfident (Jiang & Pell, 2017). Dral, Heylen, & op den Akker (2011) suggested that low speed of talking as a property of uncertainty. Moreover, wider pitch range caused listeners to perceive the utterances more certain than narrow pitch range caused. The dynamic change of pitch helped to distinguish between confidence and unconfidence. With respect to fundamental frequency range, the utterances which were evaluated as confident had highest value. Utterances with most pitch changes and faster speech rate were perceived as confident (Jiang & Pell, 2017). When locative pronouns were modified by epistemic adverb "kesinlikle" (certainly) (i.e., "kesinlikle burada", "kesinlikle şurada", and "kesinlikle orada"), perceivers relied on duration in which shorter duration was the signal for certainty, and locative pronouns did not cause speakers to be perceived as more confident.

4.11 Block IV.II

The text "Block IV.II" was displayed and stayed for 3000ms during which the participants pressed a key to continue. The switched screen displayed the specific information related to the second part of the fourth block in which locative pronouns were modified by another adverb "galiba" (probably) would be heard. The stimuli were "galiba burada" (It is probably here), "galiba orada" (It is probably there), and "galiba şurada" (It is probably there). The block was initiated by pressing spacebar. The

participants were required to listen to each audio clip by pressing spacebar (see Figure 4.3) and evaluate them on the 7-point Likert scale between 1 (not certain at all) and 4 (neutral) by considering how sure the speakers were in their answers (see Figure 4.16). The participants listened to each audio clip by pressing the spacebar and rated them on the Likert scale by clicking on one of the options. The next audio clip was presented by a click on the "next" button.

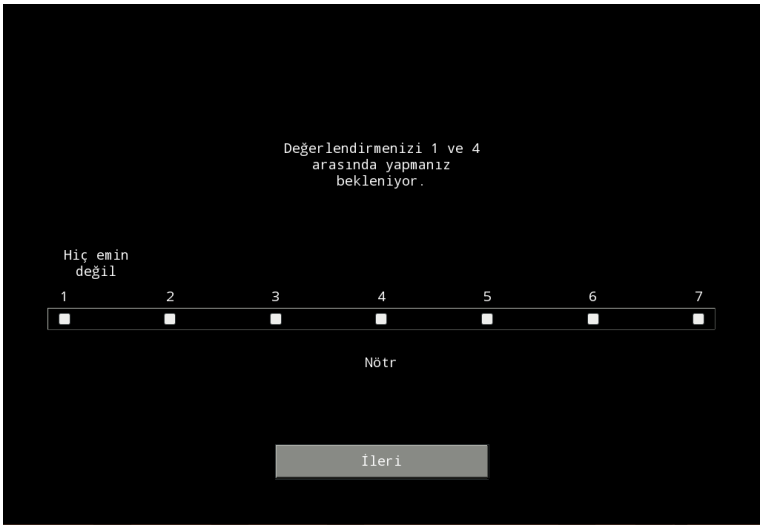

Figure 4.16: The evaluation on a 7-point Likert scale in Block IV.II

4.11.1 Data preparation & Analysis

An Excel file was created for Block 4.2 with the columns as it was prepared for Block 4.1. The participants evaluated items on a 7-point Likert scale between 1 (not certain at all) and 4 (neutral) by considering how sure the speakers were in their answers in the perception experiment. The scores were normalized according to the value of 4. Thus, the maximum uncertainty score a speaker could get was the value of 3 and the minimum uncertainty score a speaker could get was the value of 0.

All of the utterances elicited from the Experiment 1 (production study) in which for each of the 7 speakers there were three utterances (1 modality * adverb + 3 locative), for a total of 21 utterances were involved in the second part of the fourth block of the Experiment 2 (perception study). A total of 672 responses were recorded from the perception study of Block IV.II (21 responses x 32 participants). The recorded responses composed a continuous variable that is speaker confidence score (uncertainty score). Therefore, LMM was performed to be able to analyse continous data. The fixed factors (independent variables) were acoustic properties (i.e., mean pitch, duration, and pitch range), adverb + locative (with three levels: "Galiba burada" (It's probably here), "Galiba şurada" (It's probably there), and "Galiba orada" (It's probably there)), and modality (with one level: uncertain). The random factors were participant.no and item.

4.11.2 Result

At the end of the fixed and random structuring the selected model is as the following:

Formula: Speaker_Confidence_Score ~ C.Dur + (1 | Participant.no) + (1 | item)

Standard deviation is a measure of the variability for the random effects that are participants and items within the model. Items "item" (*SD* = 0.3554) had more variability than participants "participant.no" (*SD* = 0.2380).

Table 4.9: The coefficient table for the effect of duration

Fixed effects:				
	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	1.371e+00	9.340e-02	14.673	3.61e-15 ***
C.Dur	2.037e-03	3.605e-04	5.651	1.32e-05 ***

The estimate value for intercept indicates that when centered duration is zero (the value is mean duration) predicted speaker confidence score for uncertainty is 1.371e+00 (*b* = 1.371e+00, *SE* = 9.340e-02, *t* = 14.673, *p* < .001). The coefficient "C.Dur" (centered duration) is the slope for the continuous effect of duration. 2.037e-03 means that when centered duration is increased one unit, speaker confidence score for uncertainty is increased by 2.037e-03. Statistically significant result indicates that speaker confidence score for uncertainty is higher when duration is longer *(b* = 2.037e-03, *SE* = 3.605e-04, *t* = 5.651, *p* < .001). R^2 value 0.4277621, indicating that about 42.7% of variance is able to be explained by the selected model.

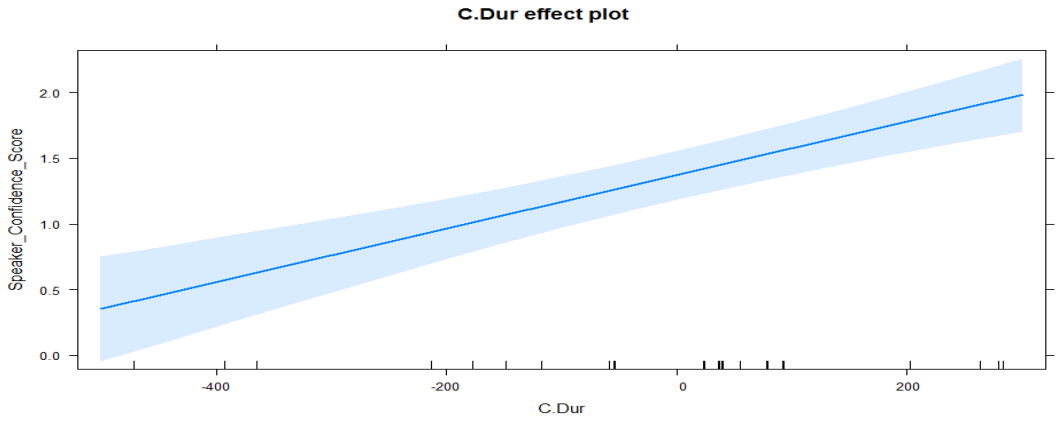

Figure 4.17: Showing the effect of duration on speaker confidence score

4.11.3 Discussion

For locative pronouns modified by epistemic adverb "galiba" (probably), listeners perceived the items more uncertain when they had longer duration. Duration of the stimuli effects how pitch contour is perceived (Chen, Zhu & Wayland, 2017). Utterances with longer duration, and slower speech rate were perceived as unconfident (Jiang & Pell, 2017). Dral, Heylen, & op den Akker (2011) suggested that low speed of talking and low intensity as properties of uncertainty. When locative pronouns were modified by epistemic adverb "galiba" (probably) (i.e., "galiba burada", "galiba şurada", and "galiba orada"), perceivers relied on duration in which longer duration was the signal for uncertainty, and locative pronouns did not change their speaker confidence score and caused speakers to be perceived as more unsure. This result might be due to lenght of vowel on the utterance "galiba" (probably). People tend to perceive "galiba" (probably) to be less certain from the long vowel (irrespective of meaning). Thus delays are signalized through duration.

4.12 Summary of the results

Locative pronouns with shorter durations determined perception of the confidence of speaker as certain whereas longer durations determined it as uncertain. Moreover, Locative pronouns with wider pitch range determined perception of the confidence of speaker as certain whereas narrower ranges determined it as uncertain. "Burada" (here) was perceived as the most certain followed by "şurada" (there), and "orada" (there). Independently of modality of the locative pronouns, lower pitch value facilitated correct estimation of the speaker confidence. The number of correct match is greater for certain modality.

Acoustic features did not have a influence on epistemic adverb "kesinlikle" (certainly) in perception of its modality. Labeled modality predicted the perceived speaker confidence. Independently of modality of epistemic adverb, high pitch value facilitated correct estimation of the speaker confidence.

For epistemic adverb "galiba" (probably), listeners perceived the items with longer durations as uncertain. Also, listeners perceived higher pitch as uncertain.

For evaluation of locative pronouns modified by epistemic adverb "kesinlikle" (certainly), listeners relied on shorter total duration for higher certainty. Also, they relied on higher pitch range to pick higher certainty.

For evaluation of locative pronouns modified by epistemic adverb "galiba" (probably), listeners relied on longer total duraton for higher uncertainty.

Because our group of speakers may not have expressed certainty/uncertainty in the same way consistently, the listeners may not have been able to extract that information from all speakers. The random variable speaker always had higher variability than the subjects. The results could just show that.

CHAPTER 5

INTONATION ANALYSIS

This chapter begins by describing clustering procedure of the utterances which were produced in the Experiment 1 based on the results of the Experiment 2. The chapter explains how wav files were annotated. Moreover, the chapter demonstrates pitch tracks in terms of speaker confidence categories, and occurences of both pitch accents and boundaries within elicited utterances among these speaker confidence categories. Speaker confidence categories were based on the participant responses from Experiment 2 rather than the confidence categories of Experiment 1. Furthermore, it attempts to explain common ground as well as discrepancies observed in intonation by providing examples from the data. In this chapter, the terms intonation contour and pitch contour are used interchangeably to mean both pitch accents and boundaries.

5.1 Clustering

The utterances "burada" (here), "şurada" (there), and "orada" (there) were clustered into three main categories based on the speaker confidence. These three main categories included "certain", "in-between", and "uncertain" categories. Beside, there existed a subcategory of "uncertain" category which was named marginally uncertain. When these categories were determined, the perceptions of the participants obtained from the Experiment 2 were taken into account. Correct scores were calculated for each utterance. The utterance had score of 1 only its labeled modality, meanly modality, (i.e., certain or uncertain) and perception of the participant, meanly perceived modality, (i.e., certain or uncertain) were matched. The term "modality" refers to the category labeled in Experiment1. The term "perceived modality" refers to the category based on the speaker confidence responses elicited from Experiment 2. Regardless of the modality of the utterance, each utterance was assigned to a category according to correct scores they had. The maximum score an utterance could get was thirty-two. A chi-square test of goodness-of-fit was performed to determine whether the two modalities were equally perceived. Perception for the modalities was not equally distributed, X^2 (1, N = 32) = 4.50, $p < .05$. Chi-square test result was significant ($p = .034$) if an utterance could get twenty-two out of thirty-two. This score corresponded to 68,75% and the value was determined as cut off. The utterances which had the value above 68,75% were assigned to a category as to their modality (i.e., either certain or uncertain). The value 65,625% was considered as marginally uncertain or marginally certain category. The utterances

that meet the requirement by having value of 65,625% were assigned to a category as to their modality as well. Since there was only three utterances with uncertain modality had this value, there was only three marginally uncertain categories. The utterances that had values between 68,75% and 31,25% were assigned to in-between category. An utterance with a value below 31,25% was assigned to the opposite category of its modality (i.e., if the modality was certain, the item was assigned to uncertain category).

5.2 Annotation

This section explains the way of annotating the utterances in the present study by a software and a labeling guide.

The Annotation analyses were conducted by using Praat software (Boersma & Weenink, 2017). In order to conduct intonation analysis, wav files were trimmed manually by moving cursor to the regions with audio signal and sound files were created through "Extract selected sound (time from 0)" command. These trimmed sounds were opened in the software as objects by "Read from file" command. "To TextGrid" command was used to create annotation text file with one internal tier to be able to annotate the utterance by dividing into syllables and one point tier to be able to label pitch accents. The sound file and TextGrid file opened together. Annotations were done by two annotaters together. Autosegmental-Metrical Model (AM) has a transcription system which is ToBI and this system allows for intonation studies (Pierrehumbert, 1980). Turkish intonation was analyzed in this framework. Annotation criterias were grounded on both a table based on Pierrehumbert's 1980 classiffication (as cited in Ladd, 1996, p. 82) and another table based on ToBI contours for Standard American English suggested by Hirschberg (2004, p.10). Moreover, annotaters practiced on the examples of Gussenhoven (2004). Once, two annotators were agree on labeling annotation file was saved by "save TextGrid as text file" command.

The following section illustrates pitch track graphs of the utterances which were clustered into the categories based on the cut off value explained above (See section 5.1).

5.3 Pitch Tracks

The section below illustrates the pitch tracks of the utterances "burada" (here), "şurada" (there), and "orada" (there) as well as the method of illustrating the pitch tracks in the graphs by using Praat software (Boersma & Weenink, 2017).

There are three main categories (i.e., certain, in-between, and uncertain) and one subcategory (i.e., marginally uncertain) which were determined according to the cut off value explained above. Since each utterance had different durations, duration of the utterances were equated by using a python code which added silence to the shorter ones to be able reach the duration of the longest one. The main reason for adding silence is to have the same scales for the pitch graphs to be able to visually compare them. This

addition of silence was not there when durations were compared. Pitch analysis were conducted by using Praat software (Boersma & Weenink, 2017). Equated sounds in terms of duration were added to the software as objects by "Read from file" command. Then, their pitch values were extracted by "To Pitch" command. The pitch floor was set as 75.0 Hz and the pitch ceiling was set as 500.0 Hz. In the end, after selection of all pitch values, different utterances belonging to the same category were depicted in the same graph by "Draw" command. Initially, pitch tracks of "burada" (here), "şurada" (there), and "orada" (there) will be presented as clustered into certain category within the same graph based on the results of the perception experiment. Then, pitch tracks of "burada" (here), "şurada" (there), and "orada" (there) will be presented respectively as clustered into certain category within the separate graphs based on the results of the perception experiment. Secondly, pitch tracks of "burada" (here), "şurada" (there), and "orada" (there) will be presented as clustered into in-between category within the same graph based on the results of the perception experiment. Then, pitch tracks of "burada" (here), "şurada" (there), and "orada" (there) will be presented respectively as clustered into in-between category within separate graphs based on the results of the perception experiment. Thirdly, pitch tracks of "burada" (here), "şurada" (there), and "orada" (there) will be presented as clustered into marginally uncertain and uncertain category within the same graph based on the results of the perception experiment. Then, pitch tracks of "burada" (here), "şurada" (there), and "orada" (there) will be presented as clustered into marginally uncertain category within the same graph based on the results of the perception experiment. Lastly, pitch tracks of "burada" (here), "şurada" (there), and "orada" (there) will be presented as clustered into uncertain category within the same graph based on the results of the perception experiment.

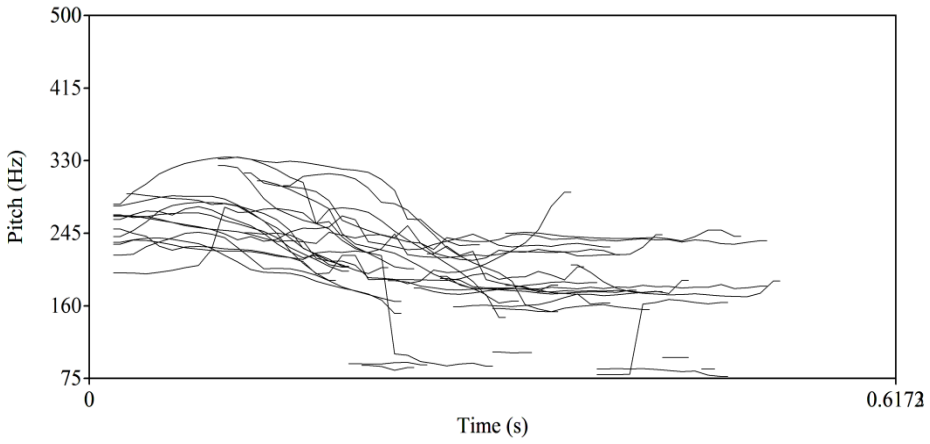

Figure 5.1: Pitch tracks of the utterances "burada" (here), "şurada" (there), and "orada" (there) in certain category

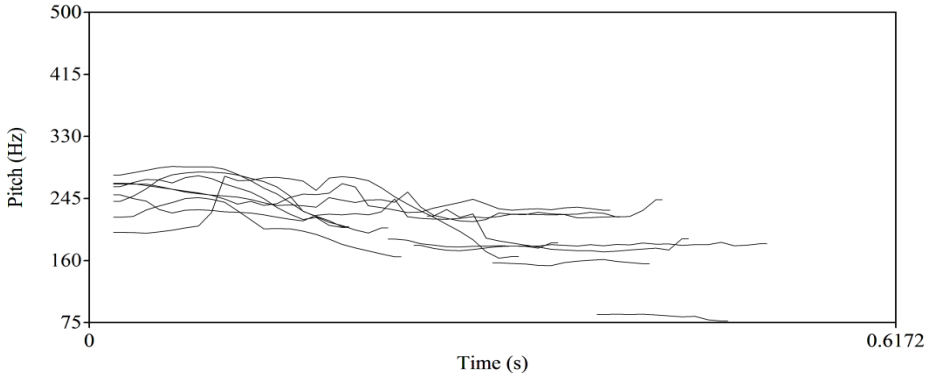

Figure 5.2: Pitch tracks of the utterance “burada” (here) in certain category

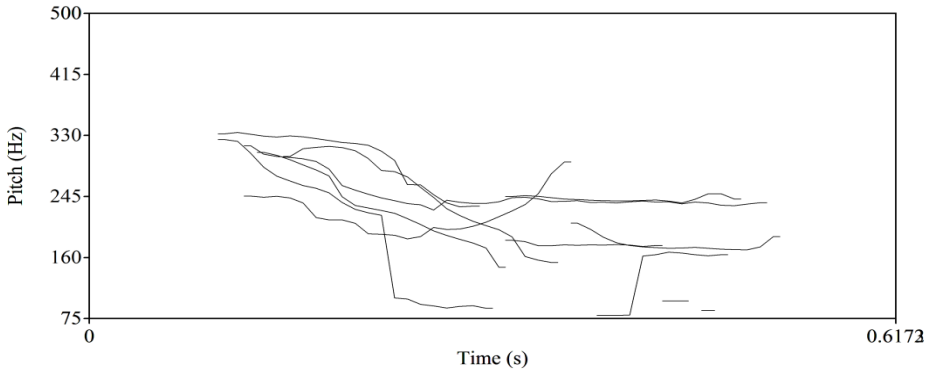

Figure 5.3: Pitch tracks of the utterance “şurada” (there) in certain category

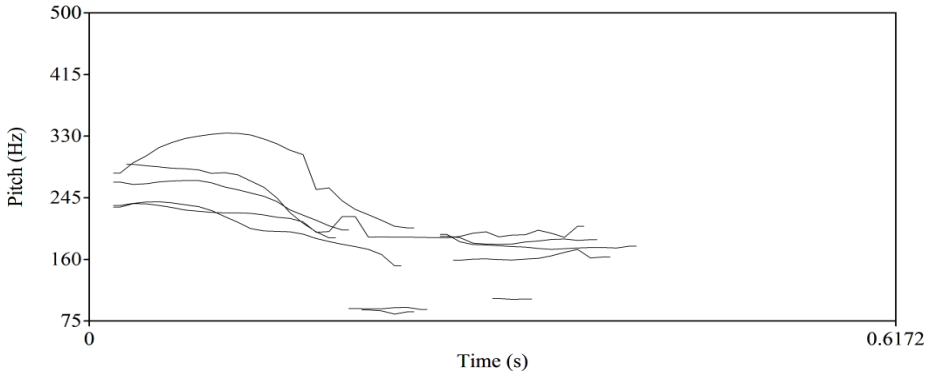

Figure 5.4: Pitch tracks of the utterance “orada” (there) in certain category

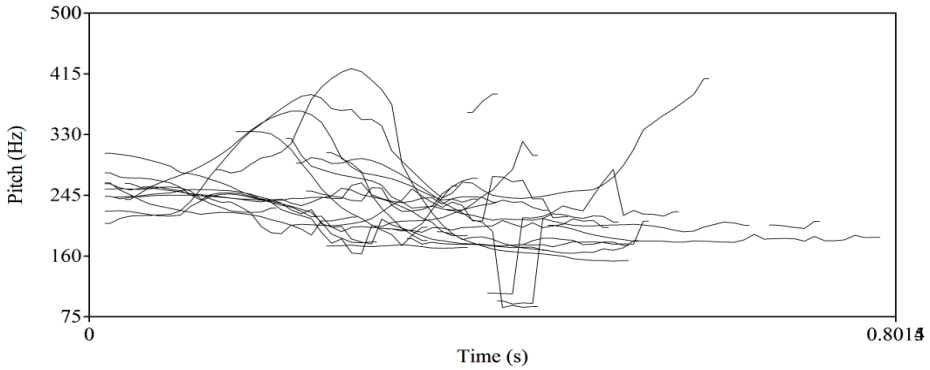

Figure 5.5: Pitch tracks of the utterances “burada” (here), “şurada” (there), and “orada” (there) in in-between category

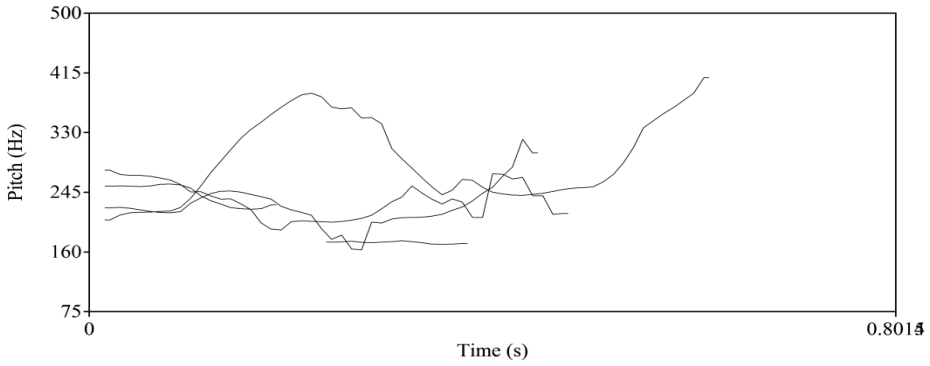

Figure 5.6: Pitch tracks of the utterance “burada” (here) in in-between category

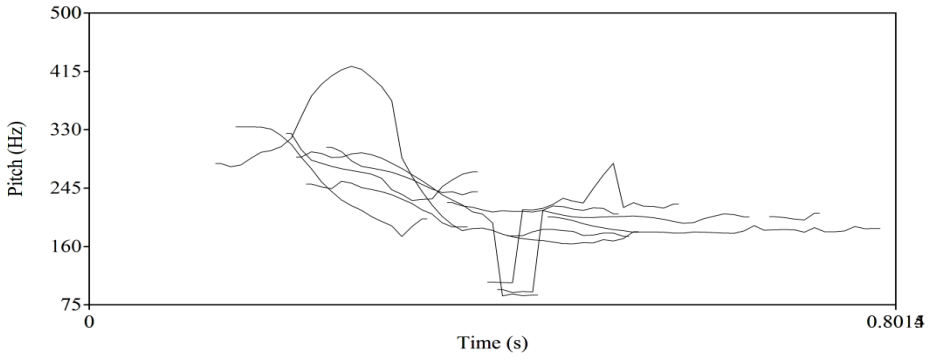

Figure 5.7: Pitch tracks of the utterance “şurada” (there) in in-between category

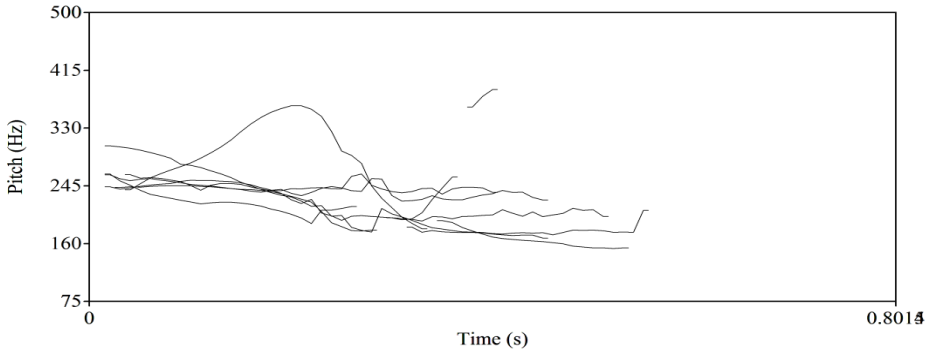

Figure 5.8: Pitch tracks of the utterance “orada” (there) in in-between category

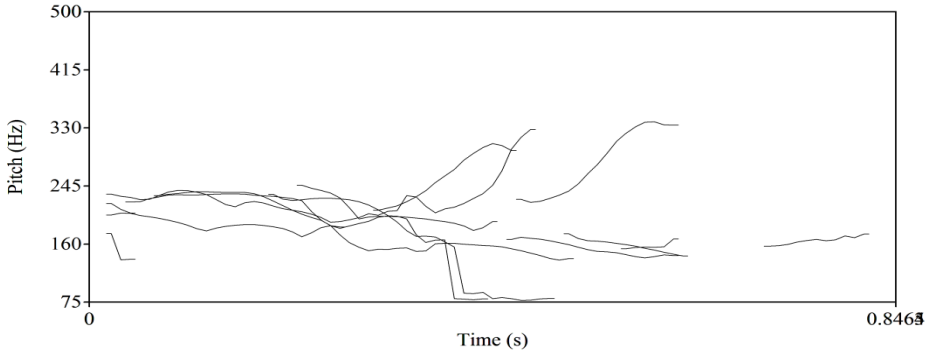

Figure 5.9: Pitch tracks of the utterances “burada” (here), “şurada” (there), and “orada” (there) in uncertain and marginally uncertain category

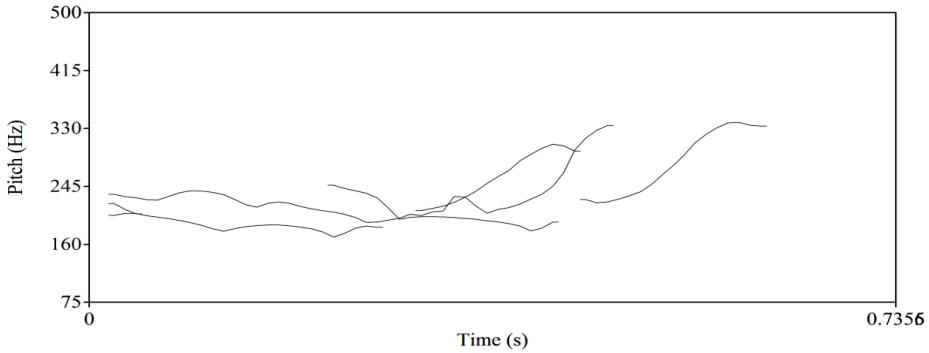

Figure 5.10: Pitch tracks of the utterances “burada” (here), “şurada” (there), and “orada” (there) in uncertain category as marginally uncertain

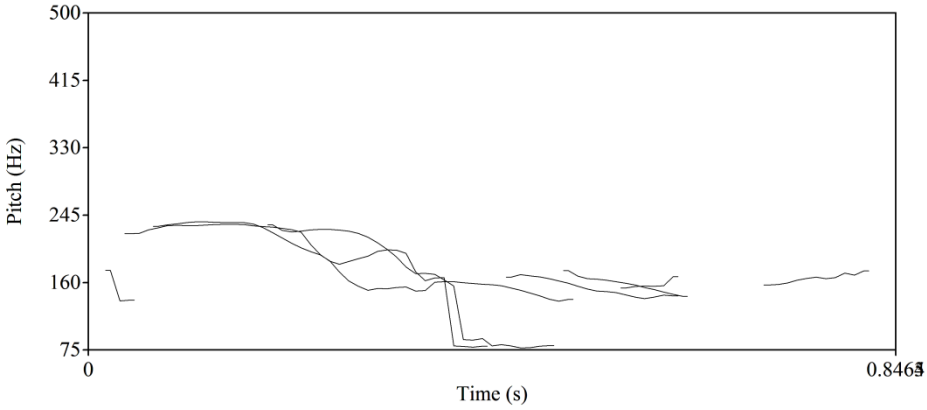

Figure 5.11: Pitch tracks of the utterances “burada” (here), “şurada” (there), and “orada” (there) in uncertain category

Overall, pitch tracks (see Figure 5.1 - 5.4) depicted a preference for falling pitch contours when expressing certainty. On the other hand, pitch tracks (see Figure 5.5 - 5.8) depicted a preference for both falling and rising pitch contours during expression. Since these utterances are categorized in in-between category, it cannot be concluded that these pitch tracks reflect features of falling and rising pitch contours precisely. In similar, pitch tracks (see Figure 5.9) depicted a preference for both falling and rising pitch contours when expressing uncertainty. Pitch tracks (see Figure 5.10) depicted rising contours in marginally uncertain category and falling contours in uncertain category (see Figure 5.11).

The following section shows the observed frequency of pitch contours among utterances voiced by the speakers in the Experiment 1 in reference to annotations of the two experimenters.

5.4 Intonation Contours

This section provides two tables. Table 5.1 presents the distribution of intonation contours belonging to the utterances by speakers in the Experiment 1 in terms of the categories (i.e., certain, in-between, uncertain) which were created with respect to the results of the perception experiment, namely Experiment 2. Moreover, Table 5.2 presents a more detailed distribution by considering the type of locative pronouns. In the remaining part of the section, the findings based on the number of occurences of the intonation contours are summarized via the tables.

Table 5.1: The number of occurrences of intonation contours within categories

Intonation Contours	**certain**	**in-between**	**uncertain**
H* LL%	6	2	0
H* HL%	2	1	0
H* LH%	1	1	0
H*+L LL%	6	6	3
H*+L HL%	1	0	0
L+H* HH%	0	1	0
L* LL%	2	4	0
L* HL%	1	0	0
L* LH%	0	1	3
L*+H LL%	0	1	0

The distribution of intonation contours attained from intonation analysis suggests that regardless of the type of locative pronouns the best predictors of certainty are H* LL% and H*+L LL% intonation contours. They are preferred over H* HL% and L* LL% intonation contours which are second within the distribution of certain category. The last and the least frequent intonation contours are H* LH%, H*+L HL%, and L* HL% within the distribution of certain category. Eventhough H* LL% is observed within the distribution of in-between category. It can be concluded that this intonation contour is more representative for certain category. However, H*+L LL%, which is another common intonation contour among certain and in-between category, is also the best predictor of in-between category. Furthermore, H*+L LL% is also observed in uncertain category and it is one of the only two predictors of this category. It should be noted that, H*+L LL% intonation contour is mutual and equally important for all categories as a predictor. This intonation contour will be approached in a more detailed way in the following section. The predictor of in-between category continues with L* LL% intonation contour which is also the second predictor of certain category yet it occurs more in in-between category than certain category. Apart from these frequent intonation contours, H* HL%, H* LH%, L+H* HH%, L* LH, and L*+H LL% intonation contours are also observed. Among these less frequent intonation contours only H* LH% is common with certain category. As for uncertain category, in addition to H*+L LL% the other predictor is L* LH%. This intonation contour can be committed to uncertain

category even if there has been one occurence of this intonation contour within in-between category. Overall, the best predictors of certainty can be established as H* LL% and H*+L LL% intonation contours. H*+L LL% and L* LL% can be dedicated to in-between category as predictive intonation contours. As regards to uncertain category, H*+L LL% and L* LH% intonation contours are the best predictors. The following section will look for the reasons behind this commonality and try to provide explanations.

Table 5.2: The number of occurrences of intonation contours within categories with respect to locative pronouns

Intonation Contours	**certain**			**in-between**			**uncertain**		
	burada	**şurada**	**orada**	**burada**	**şurada**	**orada**	**burada**	**şurada**	**orada**
H* LL%	3	0	3	0	2	0	0	0	0
H* HL%	1	1	0	1	0	0	0	0	0
H* LH%	0	1	0	1	0	0	0	0	0
H*+L LL%	2	3	1	0	2	4	1	1	1
H*+L HL%	0	1	0	0	0	0	0	0	0
L+H*HH%	0	0	0	1	0	0	0	0	0
L* LL%	1	0	1	1	1	2	0	0	0
L* HL%	1	0	0	0	0	0	0	0	0
L* LH%	0	0	0	0	0	1	1	1	1
L*+H LL%	0	0	0	0	1	0	0	0	0

The distribution of intonation contours attained from intonation analysis with respect to the locative pronouns suggests that H* LL% intonation contour which is one of the best predictors of certainty was not used with “şurada” (there). Moreover, H*+L LL% intonation contour which is the other best predictor of certainty was mostly used with “şurada” (there). However, H* LL% intonation contour is only used with “şurada” (there) in in-between category. The intonation contour H*+L LL% which is the best predictor for in-between category was mostly used with “orada” (there) and secondly

with “şurada” (there). Further, L* LL% intonation contour which is the other best predictor of in-between category was mostly used with “orada” (there). Lastly, the locative pronouns were equally used with H*+L LL% and L* LH% intonation contours which are the best predictors of uncertainty. Moreover, H*+L LL% is the intonation contour of uncertain category whereas L* LH% is the intonation contour of marginally uncertain category.

The following section provides comparative analysis of intonation contours by considering the relationship between the three main categories (i.e., certain, in-between, and uncertain), intonation contours (pitch accent + intonational boundary), duration, and pitch range.

5.5 Comparative Analysis of Intonation Contours

This section below explains how to use Praat Picture feature to make comparisons through visual graphics. Then, the best representatives of the intonation contours among the data are included in the examples. The examples are provided for each intonation contour that appeared in the data. The section is concluded with possible explanations accounted for the predictive power of intonation contours.

In order to conduct comparative analysis of intonation contours, the trimmed sounds and TextGrid files were opened in the software as objects by “Read from file” command. The stereo sounds were converted to mono by “Convert to mono” command. Then, pitch values were extracted by “To Pitch” command. The pitch floor was set as 75.0 Hz and the pitch ceiling was set as 500.0 Hz. After elicitation of these main three objects, each object was drawn to the Praat Picture separately. Praat pictures comprised a waveform, duration, pitch value, pitch track, syllabified annotation of the utterance, and intonation contour.

The examples will be given and discussion will be provided regarding similarities and dissimilarities among intonation contours.

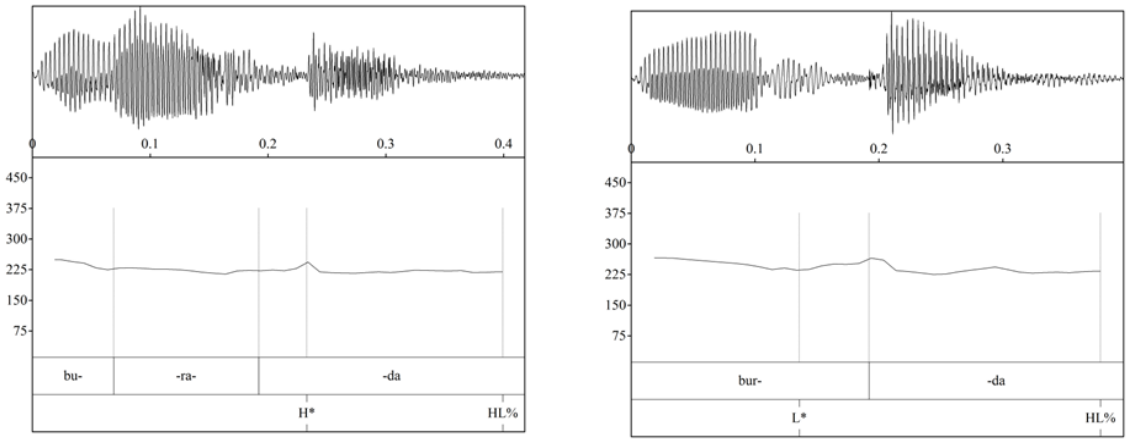

Figure 5.12: An example of L* HL% intonation contour belonging to the utterance "burada" (here) (right-hand side) and an example of H* HL% intonation contour belonging to the utterance "burada" (here) (left-hand side)

These utterances "burada" (here) on the right-hand side and "burada" (here) on the left-hand side were clustered into certain category based on the results of the perception experiment. Based on the categorization, it can be concluded that HL% intonational boundary is one indicator of certainty regardless of pitch accent. It is known by looking at the percentages that the utterance on the right-hand side with L* HL% intonation contour perceived as certain higher than the utterance on the left-hand side with H* HL% intonation contour (87,50% and 78,125%, respectively). Statistical analyses revealed a significant effect of the duration of utterance in the evaluation of speaker confidence (perceived modality). Due to the fact that these two utterances have short durations (0.4178 sec and 0.3972 sec, left-to-right), they might be perceived as certain. With respect to previous research, the accent type H* is to indicate the utterance is new and tried to be added to the belief of the addressee. With the use of L* pitch accent, the syllable becomes more "salient". The aim is not to add something to the belief between speaker and addressee (Pierrehumbert & Hirschberg, 1990, p.292). H* HL% intonation contour is employed to "elaborate" and "provide support" (Ward & Hirschberg, 1985, p. 291).

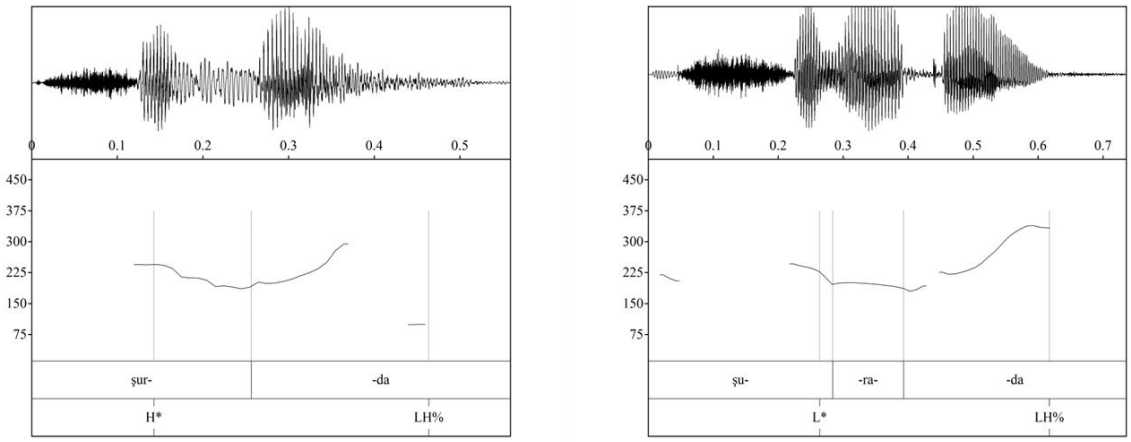

Figure 5.13: An example of L* LH% intonation contour belonging to the utterance "şurada" (there) (right-hand side) and an example of H* LH% intonation contour belonging to the utterance "şurada" (there) (left-hand side)

The utterance "şurada" (there) on the right-hand side was clustered into marginally uncertain category whereas the utterance "şurada" (there) on the left-hand side was clustered into certain category based on the results of the perception experiment. Based on the categorization, it cannot be concluded that LH% intonational boundary is one indicator of whether certainty or uncertainty. The pitch contours are differed from the pitch accents. A claim such as a start with H* pitch accent and ending with LH% intonational boundary is indicator of certainty whereas a start with L* pitch accent and ending with LH% intonational boundary is indicator of uncertainty might be weak. It is known by looking at the percentages that the utterance on the right-hand side with L* LH% intonation contour perceived as less certain (marginally uncertain) than the utterance on the left-hand side with H* LH% intonation contour (34,375% and 71,875%, respectively). It is more reasonable to rely on the effect of duration based on the significant results of the previous chapter rather than relying on a claim that a start with H* pitch accent is indicator of certainty. Eventhough pitch contours look quite similar, due to the fact that the utterance on the right-hand side with L* LH% intonation contour has longer duration (0.7356 sec) than the utterance on the left-hand side (0.5589 sec) the utterance on the left-hand side might be perceived as more certain. With respect to previous research the accent type H* is to indicate the utterance is new and tried to be added to the belief of the addressee. With the use of L* pitch accent, the syllable becomes more "salient". The aim is not to add something to the belief between speaker and addressee (Pierrehumbert & Hirschberg, 1990, p.292). H% boundary tone connects the utterance to the following one. It gives the feeling of incompleteness (Ergenç, 2002; Pierrehumbert & Hirschberg, 1990). L* LH% is employed when either addressee is already sharing the same belief or addressee is expected to share the same belief. The latter one shows "insulting effect" (Pierrehumbert & Hirschberg, 1990, p. 292).

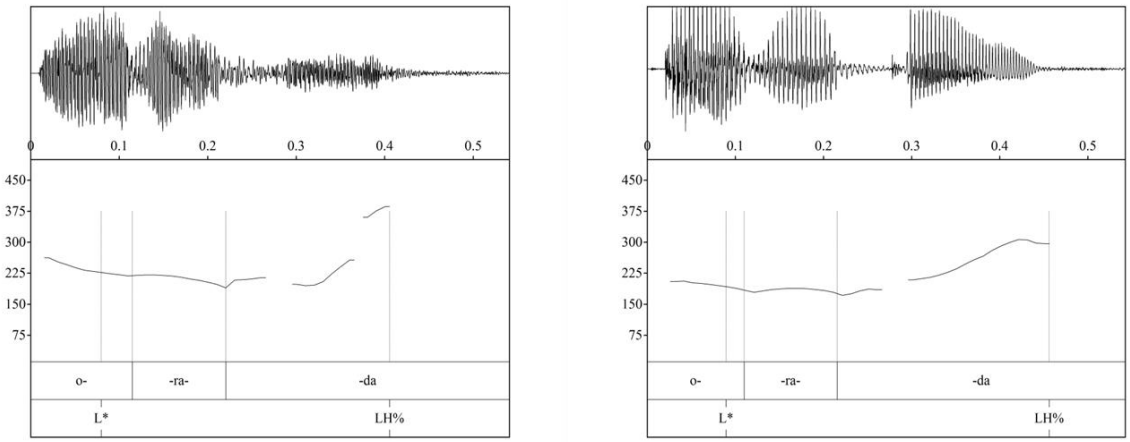

Figure 5.14: An example of L* LH% intonation contour belonging to the utterance "orada" (there) (right-hand side) and an example of L* LH% intonation contour belonging to the utterance "orada" (there) (left-hand side)

The utterance "orada" (there) on the right-hand side was clustered into marginally uncertain category whereas the utterance "orada" (there) on the left-hand side was clustered into in-between category based on the results of the perception experiment. With respect to previous research, H% boundary tone connects the utterance to the following one. It gives the feeling of incompleteness (Ergenç, 2002; Pierrehumbert & Hirschberg, 1990). Based on the categorization, it cannot be concluded that LH% intonational boundary is one indicator of whether certainty or uncertainty. The intonation contours are not differed from each other. They both started with L* pitch accent. A claim such as a start with L* pitch accent and ending with LH% intonational boundary is indicator of uncertainty might be weak. It is known by looking at the percentages that the utterance on the right-hand side with L* LH% intonation contour was perceived as less certain (marginally uncertain) than the utterance on the left-hand side with L* LH% intonation contour (in-between) (34,375% and 53,25%, respectively). It is more reasonable to rely on the effect of duration based on a significant result than relying on a claim that a start with L* pitch accent is indicator of uncertainty. However, not only intonation contours are the same but also these utterances have pretty close durations (0.5408 sec and 0.5423 sec, from left-to-right). Concluding that longer duration might be resulted in higher uncertainty might be weak here. The only suggestive claim might be possible that is considering pitch range. It is more reasonable to rely on the effect of pitch range based on the statistical analyses. Speaker confidence (perceived modality) was close to be perceived as certain when pitch range got higher. In line with this result, the utterance on the right-hand side with L* LH% intonation contour which falls into marginally uncertain category has lower pitch range and longer duration than the utterance on the left-hand side with L* LH% intonation contour which falls into in-between category (135.595 Hz and 197.3719 Hz, respectively). The effect of pitch range stands out.

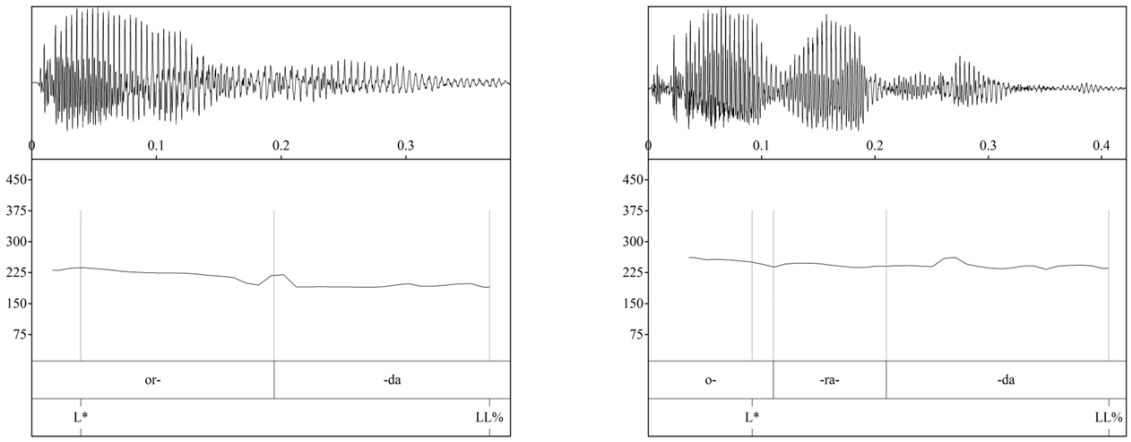

Figure 5.15: An example of L* LL% intonation contour belonging to the utterance "orada" (there) (right-hand side) and an example of L* LL% intonation contour belonging to the utterance "orada" (there) (left-hand side)

The utterance "orada" (there) on the right-hand side was clustered into in-between category whereas the utterance "orada" (there) on the left-hand side was clustered into certain category (left-hand side) based on the results of the perception experiment. Based on the categorization, it cannot be concluded that LL% intonation contour is one indicator of whether certainty or uncertainty. The intonation contours are not differed from each other. They both started with L* pitch accent. A claim such as a start with L* pitch accent and ending with LL% intonational boundary is indicator of certainty might be weak. It is known by looking at the percentages that the utterance on the right-hand side with L* LL% intonation contour was perceived as less certain (in-between category) than the utterance on the left-hand side with L* LL% intonation contour (certain category) (50% and 78,125%, respectively). It is more reasonable to rely on the effect of duration based on a significant result. Although pitch contours look quite similar, the utterance on the left-hand side has shorter duration than the utterance on right-hand side (0.3836 sec and 0.4217 sec, from left-to-right). Concluding that shorter duration might cause an utterance to be perceived as more certain might be reasonable. With respect to previous research, with the use of L* pitch accent, the syllable becomes more "salient". Speaker does not aim to change the belief of the addressee and is not certain. However, in our example boundary tone LL% does not need to be construe with the following utterance. As the falling at the end of an utterance increases, the meaning which is conveyed as completeness becomes more complete (Pierrehumbert & Hirschberg, 1990).

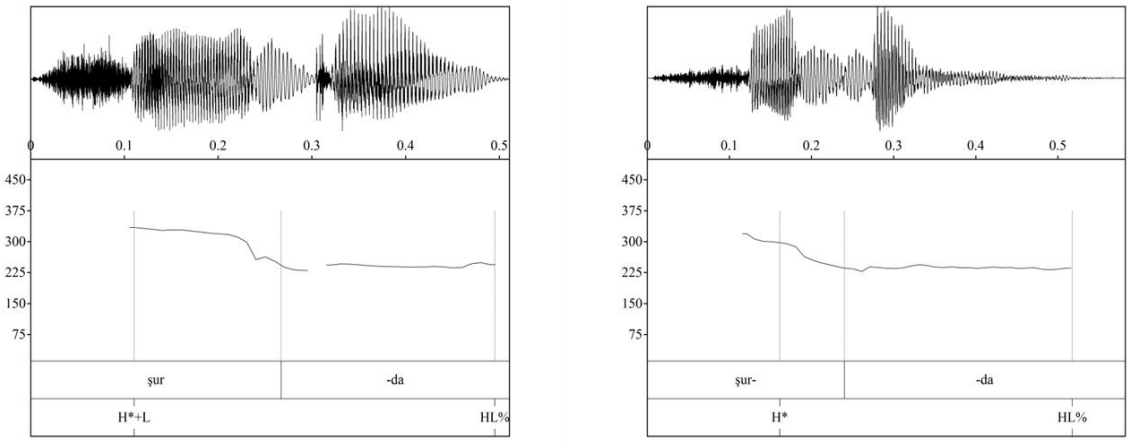

Figure 5.16: An example of H* HL% intonation contour belonging to the utterance "şurada" (there) (right-hand side) and an example of H*+L HL% intonation contour belonging to the utterance "şurada" (there) (left-hand side)

The utterances "şurada" (there) on the right-hand side and "şurada" (there) on the left-hand side were clustered into certain category based on the results of the perception experiment. Based on the categorization, it can be concluded that HL% intonational boundary is one indicator of certainty regardless of pitch accent. However, we do not know that whether it purely depends on the short duration of the utterances. In this scenario, we have only H tone as a start. The utterance on the right-hand side with H* HL% intonation contour has higher certainty than the utterance on the left-hand side with H*+L HL% intonation contour (90,625% and 78,125%, respectively). However, the utterance with the shorter duration is not the one with higher certainty (0.5819 sec and 0.5105 sec, from left-to-right). Eventhough H*+L HL% intonation contour has shorter duration, it is not more certain than the H* HL% intonation contour. The only difference apart from duration between intonation contours of these utterances is L tone. Since we could not observe the overriding effect of duration, this brings into mind that L tone might have a regressive effect to characterize more tense. However, duration still might be the reason for these utterances to be grouped under certain category. It is more reasonable to rely on the effect of duration based on a significant result. Eventhough pitch contours differed, having short durations (0.5819 sec and 0.5105 sec, from left-to-right) might have resulted in certain category. With respect to previous research, H* HL% intonation contour is employed to "elaborate" and "provide support" (Ward & Hirschberg, 1985, p. 291) whereas H*+L HL% intonation contour is employed as "calling contour" (Pierrehumbert & Hirschberg, 1990, p. 299).

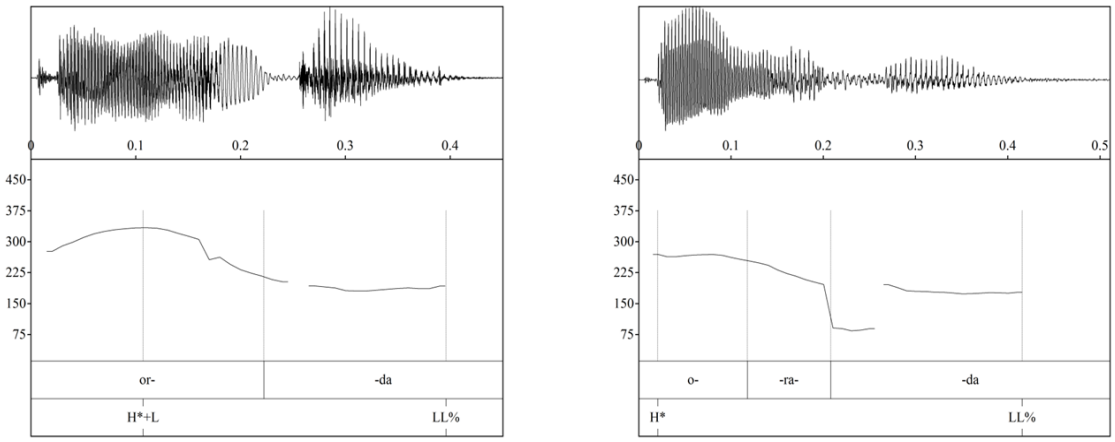

Figure 5.17: An example of H* LL% intonation contour belonging to the utterance "orada" (there) (right-hand side) and an example of H*+L LL% intonation contour belonging to the utterance "orada" (there) (left-hand side)

The utterances "orada" (there) on the right-hand side and "orada" (there) on the left-hand side were clustered into certain category based on the results of the perception experiment. Based on the categorization, it can be concluded that LL% intonational boundary is one indicator of certainty. We might want to know if this is caused by a start with H tone. The intonation contours are differed from only that the utterance on the left-hand side has L tone. We used H*+L LL% intonation contour to characterize more tense rather than H* LL% intonation contour which is characteristic of declarative contour. It is known by looking at the percentages that the utterance on the right-hand side with H* LL% intonation contour has higher certainty value (93,75%) and higher duration (0.5108 sec). On the other hand, the utterance on the left-hand side with H*+L LL% intonation contour has lower values for both certainty (81,125%) and duration (0.4502 sec). These two features are in contrast. This time duration is not able to predict certainty of the items. It is possible to say there exist an effect of L tone for the utterance on the left-hand side with H*+L LL% intonation contour. However, duration still might be the reason for these utterances to be grouped in the certain category. It is more reasonable to rely on the effect of duration based on a significant result. Eventhough intonation contours differed, having short durations (0.4502 sec and 0.5108 sec, from left-to-right) might have perceived in certain category. With respect to previous ressearch, H*+L LL% intonation contour is employed when "reading instructions" (Pierrehumbert & Hirschberg, 1990, p. 299). Moreover, H*+L intonatin contour is employed in the aim of the speaker to change the belief of the addressee as in H* accented tone (Pierrehumbert & Hirschberg, 1990).

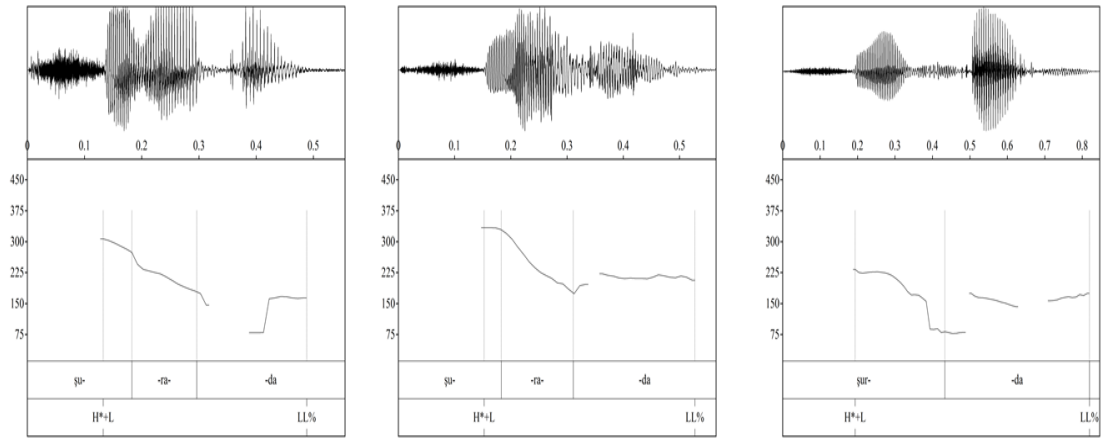

Figure 5.18: An example of H*+L LL% intonation contour belonging to the utterance "şurada" (there) (right-hand side), an example of H*+L LL% intonation contour belonging to the utterance "şurada" (there) (in the middle), and an example of H*+L LL% intonation contour belonging to the utterance "şurada" (there) (left-hand side)

The utterance "şurada" (there) on the right-hand side was clustered into uncertain category, the utterance "şurada" (there) in the middle was clustered into in-between category, and the utterance "şurada" (there) on the left-hand side was clustered into certain category based on the results of the perception experiment. Based on the categorization, it cannot be concluded that LL% intonational boundary is whether an indicator of certainty or uncertainty. In the previous example, we observed that even if L tone had a regressive effect on certainty, the utterances fell into still certain category. In the same example, the predictor was having shorter durations regardless of intonation contour (either H* or H*+L). We also observed that L* LL% intonation contour might fall into categories such as in-between and certain with respect to duration of the utterance. In the present example, we might foreseen that a start with H*+L pitch accent will result in certain category if we ignore the duration factor. In this example, we would like to know whether durations of the utterances are in parallel with our assumption. Firstly, we will compare certainty of the utterances. The utterance on the right-hand side has the highest certainty (96,875%). The utterance in the middle comes the second one in the order (53,125%), and the utterance on the left-hand side comes the third one in the order (9,375%). In terms of duration the utterance on the left has the shortest duration (0.5551 sec). As in the certainty order, in terms of duration the utterance in the middle comes the second (0.5649), and the utterance on the left-hand side comes the third (0.8464). All of these values are in line with our expectation that is regardless of intonation contours the effect of duration is ascertained.

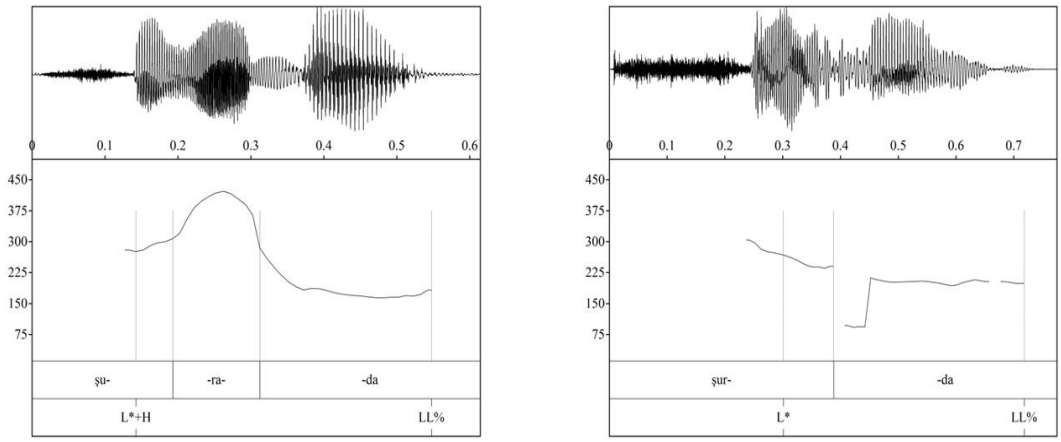

Figure 5.19: An example of L* LL% intonation contour belonging to the utterance "şurada" (there) (right-hand side) and an example of L*+H LL% intonation contour belonging to the utterance "şurada" (there) (left-hand side)

The utterances "şurada" (there) on the right-hand side and "şurada" (there) on the left-hand side were clustered into in-between category based on the results of the perception experiment. Based on the categorization, it cannot be concluded that LL% intonational boundary is an indicator of certainty or uncertainty. The intonation contours are differed from that only the utterance on the left-hand side has H tone. Although patterns of the pitch tracks are quite different it can be inferred by looking at the percentages that the utterances have equal values of both certainty (43,75%) and uncertainty (56,25%). It can be inferred from the figures as well that the utterance on the right-hand side with L* LL% intonation contour has longer duration. In the present example duration is not able to distinguish the items from each other. However, duration still might be the reason for them to be grouped under in-between category. It is more reasonable to rely on the effect of duration based on a significant result. In addition to duration, the effect of pitch range was also observed to predict speaker confidence (perceived modality) categories. Eventhough pitch contours differed, having longer durations (0.6142 sec and 0.7745, left-to-right) might have resulted with higher uncertainty and to be grouped under in-between category. With respect to previous research, intonation contours such as L*+H LH%, L*+H HH%, and L*+H HL% are described as signaling speaker uncertainty due to L*+H pitch accent based on the work of Ward & Hirschberg (1985). Our example has LL% boundary which might cause the utterance result in in-between category. Since, as the falling at the end of an utterance increases, the meaning which is conveyed as completeness becomes more complete (Pierrehumbert & Hirschberg, 1990).

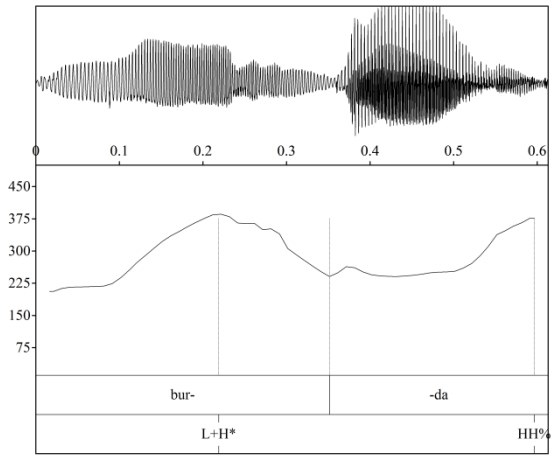

Figure 5.20: An example of L+H* HH% intonation contour belonging to the utterance "burada" (here)

This is a unique example in the data. It is the only item which has an ending with HH% boundary tone. As this item was grouped under in-between category based on the results of the perception experiment, it is not possible to foresee the predictiveness of this contour whether it indicates certainty or uncertainty. However, since HH% boundary tone gives the feeling of incompleteness, it resulted in the categorization of in-between. L+H* pitch accent aims to change belief of the addressee due to H* accented tone. In here, we should also state certainty and duration values of the utterance (53,125% and 0.6129 sec, respectively).We need more data to be able to make more reliable inferences regarding this intonation contour. With respect to previous research the most clause description is that L+H* LH% intonation contour is observed when marking "correction" and "contrast" (Pierrehumbert & Hirschberg, 1990, p. 296). Eventhough H* accented tone aims changing belief of the addresse and indicates certainty, HH% boundary tone which is employed in yes-no questions gives the feeling of uncertainty (Pierrehumbert & Hirschberg, 1990). Thus, the utterance might result in in-between category.

To conclude, the examples from the data which were inspected in this section allow us to make inferences with respect to intonation contours. With respect to table H* accented tone is used for certainty whereas L* accented tone is used for uncertainty. Particulary, the ending intonational boundary has been emphasized. Within the cases in which the ending intonational boundary is HL%, the category of the utterance is more likely to be certain. Another example for ending with HL% intonatonal boundary could be that H is the starting tone. Then, L tone is added. In our examples provided by the data, we observed that adding L tone did not cause to change the category of the utterance.

However, even the utterance with H*+L HL% intonation contour has shorter duration, this effect was surpassed by L tone. Within the cases in which the ending intonational boundary is LH% the category of the utterance is certain if H tone is the start wheras the category of the utterance is uncertain if the start is L tone. For the cases L tone is the start and the ending is LH% intonational boundary if the duration is shorter there is a decrease in uncertainty and the likelihood approaches to being grouped under in-between category. LH% intonational boundary is associated with questions. However, in the example provided by the data having a question intonation did not lead to certainty or uncertainty on its own. LL% intonational boundary was observed within all categories. For instance, the ending is LL% intonational boundary yet due to duration utterances falls into in-between category. However, the effect of duration is surpassed in some cases. Considering a certain category the ending is with LL% intonational boundary and starting with H tone. Adding L tone might have resulted with a decrease in certainty and surpassed the increasing effect of short duration in certainty. It should be noted that the duration is a significant predictor and pitch range might have a contribution to predict the speaker confidence (perceived modality) category. As to properties of the intonational curves we can work out more precisely on predictiveness of certainty and uncertainty. Based on what our data provided, falling pitch contours can be preferred while expressing certainty. It is also observed that, there is a preference for both falling and rising pitch contours when expressing uncertainty. Falling contours were found in uncertain category whereas rising contours were found in marginally uncertain category.

In this chapter, the method of clustering was explained. After the categorization of the items, utterances were depicted in the graphs in terms of pitch tracks by using Praat sotware. During depiction, both category of the items and the type of locative pronouns were considered. How to annotate in Praat was tried to be explained and also labeling guides were referenced (Gussenhoven, 2004; Hirschberg, 2004; Ladd, 1996). After labeling intonation contours in the data, the distribution of the contour types were presented in the tables by considering both categories they fall into and the locatives they belong to. Lastly, by comparative analysis the most frequent intonation contours in the data were examined. The comparisons were made by taken into consideration pitch accents at the start, ending intonational boundaries, and fixed factors which are duration and pitch range.

CHAPTER 6

DISCUSSION

The aim of the current study was to find out which acoustic properties distinguish between epistemic (or speaker) confidence (i.e., certainty and uncertainty). To achieve that aim, production and perception studies were conducted. In these studies, only locative pronouns, adverbs, and locative pronouns modified by adverbs were uttered and evaluated. Among acoustic properties duration, mean pitch, and pitch range were investigated . Longer duration signaled uncertainty whereas shorter duration signaled certainty. Moreover, higher mean pitch signaled certainty. However, pitch range did not appear as much predictive compared to duration and mean pitch. Moreover, intonation contours of locative pronouns were examined based on Pierrehumbert (1980)'s Autosegmental-Metrical Model (AM). Thus characteristic pitch accents and boundary tones were revealed as regards speaker confidence.

This study was conducted in order to determine the effect of non-verbal and verbal cues in encoding and decoding of epistemic confidence. Our results revealed that verbal and non-verbal cues are encoded to express speaker confidence and also applied during decoding. The results suggested that epistemic confidence is conveyed by means of specific adverbs, locative pronouns, and acoustic features. Taken together, production and perception experiments have shown theacoustic features that Turkish listeners/speakers make use of to distinguish between certainty and uncertainty. The same features are used in production and perception, that is, that these two modes of processing access the same information. This might be important in view of our term "mental imagination". This tight coupling is in line with a model of simulation such that the listener is simulating not only the anticipated content of the speaker's utterance (Pickering & Garrod, 2007), but also the anticipated prosody. Such a model assumes that production and perception make use of the same linguistic codes, extracted from bottom up acoustic features.

In the next sections, effects of duration, pitch, pitch range, modality, and intonation contours will be discussed, respectively.

6.1 Effect of Duration

Two things need to be distinguished here that are duration and certainty. Since the words "burada", "şurada", and "orada" have different pronounciation times, it might be that "burada" is the shortest. This could have tested by the pronounciation times of these 3 locatives when they were pronounced with neutral intonation. According to perception of the speaker confidence, the locative pronouns with shorter durations determined perception of the confidence of speaker as certain whereas longer durations determined it as uncertain. This result suggests that uttering in short durations cause speaker to be perceived as sure whereas uttering in long durations cause speaker to be perceived as unsure. Participants tended to use low intonation when replying to questions with a non-answer such as 'I do not know' since they did not need to think metacognitively for searching the answer (Geluykens, 1987). These participants were sure about what they do not know (Krahmer & Swerts, 2005). Irrispective of intonation, the result might indicate that if the speaker has knowledge of the answer, cognitive accessibility to this answer would be quicker than it is for lack of knowledge. Not only perceivers chose duration to distinguish between certainty and uncertainty. When speakers were instructed to utter as they are unsure, they had a tendency to utter the locative pronouns longer as compared to when they were instructed to utter as they are sure. The manipulation of duration suggested that production study is compatible with perception study. The same features are used in production and perception, that is, that these two modes of processing access the same information. Production and perception make use of the same linguistic codes, extracted from bottom-up acoustic features. In addition, another result indicated that locative "burada" (here) mostly perceived as certain by the participants. These two results are in line since the locative with the shortest duration had the biggest number of certainty regarding speaker confidence. The reason could be that both speakers uttered the locative "burada" (here) in shortest duration to convey their confidence as sure and perceivers expected the locative "burada" (here) to be a signal for certainty. Even if locative "şurada" (there) has the longest duration, it had less number of certainty compared to locative "orada" (there) regarding speaker confidence. The results might indicate that the meaning of the lexical item overrides the effect of duration. A perceiver would rather expect a speaker who is certain of the location of the key to use "burada" than "surada", irrespective of the distance. We cannot make the same point for the producer, since producers were instructed by the experimenter which locative to utter. Perceivers relied on the locative pronouns rather than their durations and inferred from the locative pronouns based on distance-proximity relationship. If we consider the locative pronouns as physical representations of the hidden object according to the imagined scenario, the locative pronoun "burada" (here) which is related to the closest distance caused speakers to be perceived as the most sure by the speakers. This finding is also supported by construal level theory (Trope & Liberman, 2010) and cognitive accessibility (Burenhult, 2003; Kahneman, 2003). With respect to construal level theory psychological distance refers to epistemic distance to mental presentation of an event or object. "Construal level theory" claims that people construe events differently in terms of their spatio-temporal distance from themselves. It could thus be that distance is also sensitive to "certainty" with proximal objects (burada) being construed as more certain than distal objects (şurada, orada). Moreover, cognitive accessibility is reachability to mental content, in other words retrieval of knowledge

from memory. These two concepts are employed here to answer the question of whether space and semantic encodings are in line. We wanted to know the relationship between semantic spatial language and spatial cognition. Research question was whether language-specific encodings of spatial concepts affect nonlinguistic spatial cognition. In languages such as Jahai proximals are employed whereas in Dutch distals are employed. Since it is known that semantic spatial encoding of "bu" takes into account the distance of the referent thus used for proximality (Küntay & Özyürek, 2006), in Turkish proximals are employed. Some languages can mimic Turkish with regard to choice of demonstrative. However, since distals are employed in Dutch, we might not suggest a universal cognitive principle to explain the use of demonstratives such as detemining factor. According to collocation of these locative pronouns the order is burada (here), şurada (there), and orada (there) in terms of distance-proximity relationship (Kornfilt, 1997). It is likely that participants perceived the items as certain when they are at the closest distance to the speaker.

In a similar way, according to perception of the speaker confidence eventhough it could not be explained by the acoustic properties when adverb "kesinlikle" was uttered by the speakers, they had a tendency to utter this adverb in shorter durations to convey certainty. Moreover, according to duration when adverb "galiba" was uttered in longer durations perceivers predicted confidence of the speaker more likely to be uncertain. When speakers were instructed to utter the adverb "galiba" as they were unsure, they had a tendency to utter the token within longer durations as compared to when they were instructed to utter as they were neutral. This longer duration might be due to length of vowels and delays and these are signaled through duration. With respect to pauses, the utterances which were evaluated as unconfident had frequent pauses (Jiang & Pell, 2017).

Furthermore, according to speaker confidence degrees when locative pronouns were modified by an adverb (i.e., "kesinlikle burada", "kesinlikle şurada", "kesinlikle orada"), perceivers relied on duration. If the speakers uttered the tokens in shorter durations they were perceived more certain than they uttered in longer durations. The difference that was rooted from locative pronouns due to their semantics was overridden by the epistemic adverb "kesinlikle". Locative pronouns did not affect speaker confidence degree and did not cause speakers to be perceived as more sure. When locative pronouns were modified by another adverb (i.e., "galiba burada", "galiba şurada", and "galiba orada"), perceivers relied on duration in which longer duration was the signal for uncertainty, and locative pronouns did not cause speakers to be perceived as more unsure. Again the effect of epistemic adverb was found. This result might be due to lenght of vowels while uttering adverb "galiba". Moreover there might be pauses between adverb and locative pronoun during articulation. Since, instructions requested from speaker to utter adverb "galiba" with uncertain stance whereas uttering locative pronouns in neutral way. These pauses and delays are signaled through duration. With respect to pauses, the utterances which were evaluated as unconfident had frequent pauses (Jiang & Pell, 2017). Furthermore, if there is emphasize on the adverb "galiba" it is also signaled through duration (Pierrehumbert & Hirschberg, 1990).

6.2 Effect of Pitch

According to correct estimation of the locative pronouns whether they were uttered when speaker was sure or not, correctness accuracy is fewer when mean pitch increases. This result suggests that when speakers used their pitch in lower frequencies they were able to convey better their confidence whether certain or uncertain. Low pitch conveyed modality better. It is known that low mean pitch value has functions such as confidence, certainty, and finality (Bolinger, 1983; Balog & Brentari, 2008; Cruttenden, 1981). In order to explain the prominence of low pitch in facilitating speaker's task could be a topic for further research. Moreover, according to correct estimation of the adverb "kesinlikle" whether speaker was sure or neutral, correctnes is more (accuracy increases) when mean pitch increases. This result suggests that when speakers used their pitch in higher frequencies they were able to convey better their confidence whether certain or neutral. High pitch conveyed modality better. This might be due to stress on the words that resulted in increase in frequency and high-pitched voice (Demircan, 2001). Loudness and high pitch are considered relevance of stress by Kornfilt (1997). It is also known that high pitch functions as indicator of attention drawing, unconfidence, and continuity (Jiang & Pell, 2017; Rodero, 2011; Cruttenden, 1981). English intonation differs from that of Turkish since Ipek & Jun (2013) stated that rising pitch is employed to convey certainty. The function of high pitch with respect to conveying model could be further investigated.

Furthermore, according to perception of the speaker confidence when adverb "galiba" was uttered listeners relied on mean pitch. If mean pitch was high the speakers were perceived as more uncertain. Loudness and high pitch are considered relevance of stress by Kornfilt (1997). However, when speakers were instructed to utter within neutral condition, they have a tendency to utter the token "galiba" with higher pitch as compared to when they were instructed to utter within uncertain condition. In addition, the predictiveness of duration might depend on the semantics of the epistemic adverb in other words the relationship between semantics and pitch might not be straightforward.

6.3 Effect of Pitch Range

According to perception of the speaker confidence when locative pronouns were uttered, high pitch range resulted with perception of the confidence of the speaker as certain. When speakers utter in high pitch range, they were perceived as sure of their answers. High pitch range might be due to stress and emphasizing on a syllable more might be perceived as certain (Pierrehumbert & Hirschberg, 1990; Kornfilt, 1997; Demircan, 2001). Manipulation of pitch range can change semantic category of a perceived utterance such as friendliness, confidence, and surprise according to a perception study in English and Dutch by Chen, Gussenhoven, & Rietveld (2004). Moreover, regarding production, when speaker were requested to utter locative pronouns as certain they had a tendency to raise their pitch range. Moreover, according to speaker confidence degrees when locative pronouns were modified by an adverb (i.e., "kesinlikle burada", "kesinlikle şurada", "kesinlikle orada"), perceivers relied on pitch range, fundamental

frequency range, the utterances which were evaluated as confident had highest value (Jiang & Pell, 2017). If the speakers uttered the tokens in higher pitch range they were perceived surer than they uttered in lower pitch range. This result could be due to relation between pitch range and the adverb. The aim of speakers for manipulating their pitch range (increase or decrease in pitch range) might be to emphasize on the utterance (Pierrehumbert & Hirschberg, 1990).

Pitch range is the variable that is most strongly related to the dynamic structure of the utterance, i.e., how, over time, the prosodic contour changes. Mean pitch and mean duration, however are static or summary variables that do not tell us anything about the dynamic contour. The reason that pitch range is not very strongly predictive in Turkish might be due to Turkish being a non-tonal language. In a study by Jongman, Qin, Zhang, & Sereno (2017) perception of English and Mandarin listeners were compared in terms of lexical tones. It was found that since English is a non-tonal language perceivers from this language used pitch height (average pitch) as a cue rather than perceivers from Mandarin language used pitch slope (pitch change). The study emphasizes on language background. Another study by Gandour & Harshman (1978) suggested that in a tone language among the static parameters such as average pitch, length, and pitch slope, pitch slope is the one that conveys linguistic information. In our study average pitch was most likely to be turned out as a predictive parameter. However, it might still be that listeners of non-tonal language make use of dynamical features of language production. Since it is true for them as well that predictions should be nurished by such dynamical changes. If we can guess the pitch contour of the entire utterance this may help us in our initial guess what kind of information will be conveyed in the remainder of the utterance. This cue would even be stronger for speakers of a tonal language. This calls for future studies as well.

6.4 Effect of Modality

Correct estimation of the locative pronouns whether they are uttered in certain or uncertain modality is higher for certain modality than uncertain modality. Certain modalities were perceived more accurate. Is it easier to estimate correctly certain modality? The result shows that uncertain modalities were perceived as certain rather than uncertain. This result could be due to the ability of speakers to express their speaker confidence as certain was better than uncertain. Finally, we observed that speakers were better at expressing certainty than uncertainty based on the perception of the listeners that rated the items mostly as certain. This might be due to that speakers had an imaginary scenario and created their own cognitive context. As regards self-presentation our speakers more accurately conveyed whether they were sure of the answer rather than unsure. Krahmer & Swerts (2005) in their experiment concluded the reason that children cannot explain uncertainty is due to self-presentation. In our experiments, speakers might be less cared for self-presentation as well. Our results also seem to indicate that in communication it is more important to convey certainty than uncertainty. This might be the case because if some information is certain then we should act on it (rather than do nothing). When some information is uncertain, we should probably not act on it.

Moreover, when adverb "kesinlikle" perceived to be able decide speaker confidence perceivers relied on modality. This result shows that modality and perceived modality were in line. Speaker confidence which is certain perceived as certain significantly more than speaker confidence neutral. We cannot conclude the effect of acoustic properties on the perception, however, when speakers were instructed to utter within certain condition, they have a tendency to utter the token "kesinlikle" with higher pitch, higher pitch range, and shorter durations as compared to when they were instructed to utter within neutral condition. These findings are in line with the previous research. Since English intonation differs from that of Turkish since Ipek & Jun (2013) stated that rising pitch is employed to convey certainty. Moreover, with respect to fundamental frequency range, the utterances which were evaluated as confident had highest value. Also, faster speech rate were perceived as confident (Jiang & Pell, 2017).

6.5 Intonation Contours

Among intonation contours H*+L LL% is the most common one and observed in all categories. It is one of the best representer intonation contours for all categories. H*+L LL% intonation contour is employed when "reading instructions" (Pierrehumbert & Hirschberg, 1990, p. 299). Since it starts with H* accent tone to indicate the utterance is new and tried to be added to the belief of the addressee, it should represent certainty. When using H*+L the aim of the speaker is to change the belief of the addressee (Pierrehumbert & Hirschberg, 1990, p.292). Also, LL% boundary tone which signals the completeness of the utterance thus of message is the other cue for H*+L LL% intonation contour to be perceived as certain (Pierrehumbert & Hirschberg, 1990). Falling intonation is accompanied by declaratives and commands thus carry certainty (Bolinger, 1983; Balog & Brentari, 2008). With respect to our observations from data pitch accent affected category of the perceived modality. H* LL% was perceived as certain whereas L* LL% was perceived as in in-between category. In our study duration has a significant effect. This result might be due to both speakers had tendency to behave as they were reading instuctions and to perceive shorter durations as certain whereas longer durations as uncertain. It is known that duration of the stimuli effects how pitch contour is perceived (Chen, Zhu & Wayland, 2017). The second intonation contour to represent certainty is H* LL%. In here as well, the aim of the speaker is to change the belief of the addressee due to accented tone H* (Pierrehumbert & Hirschberg, 1990, p.292). In addition, LL% boundary tone signals the completeness as well as certainty due to falling pitch contour (Bolinger, 1983; Balog & Brentari, 2008). It is the mostly known example of downstepped contour and have declarative message (Pierrehumbert & Hirschberg, 1990; Gravano, Benus, Hirschberg, German, & Ward, 2008). Moreover, L* LL% intonation contour is the best representer of in-between category. This might be due to L* accent tone in which the speaker does not aim to change belief of the addressee and is not certain. However, answers are given with LL% boundary tone. An utterance with L% bounday tone does not need to be construe with the following utterance (Pierrehumbert & Hirschberg, 1990). The best representer of uncertainty is L* LH%. Accent tone L* does not have an aim to change the belief. Both L* and LH% represent uncertainty. Although LH% boundary tone is used for questions, it does not prevent

certainty. H* pitch accent is more effective than boundary tone. Moreover, LH% boundary tone alone does not require a reply; however, it is construe with the following utterance with LL% boundary tone (Pierrehumbert & Hirschberg, 1990, p. 308). L* LH% is employed when either addressee is already sharing the same belief or addressee is expected to share the same belief in deed addressee does not. The latter one shows "insulting effect" (Pierrehumbert & Hirschberg, 1990, p. 292). For instance, if we compare H* LL% and L* LL% intonation contours, we observed that pitch accent affected category of the perceived modality. Furthemore, HL% intonational boundary is one indicator of certainty regardless of pitch accent. Within the cases in which the ending intonational boundary is HL%, the category of the utterance is more likely to be certain.

CHAPTER 7

CONCLUSION

This chapter summarizes results of the study, discusses both conbutions and limitations of the study, and proposes future work.

7.1 Summary of the Findings

The present study was inspired by the work of Hübscher, Esteve-Gibert, Igualada, & Prieto (2017) and Jiang & Pell (2017). We wanted to answer questions such as "How do speakers convey epistemic confidence?" and "What would be the role of speech in the evaluation of epistemic confidence?" Under the light of previous findings regarding the topic, we aim to investigate the phonological, lexical and spatial dimensions of indication and perception of speaker confidence. For this purpose we used Turkish as a case domain. For the investigation of phonological aspects we used the tonal analyses of recorded elicited utterances (mean pitch, pitch range, intonational contours); for the investigation of lexical aspects, we used different epistemic adverbials (kesinlikle 'certainly', galiba 'probably'); and for the investigation of spatial factors we used demonstrative pronouns forming a scale of proximity (burada 'here', şurada 'there', orada 'there'). We were also interested in the interaction of these potential components.

Among acoustic properties the effect of duration was found to be prominent. Shorter durations yielded speakers to be evaluated as they were certain (sure of their answer) whereas longer durations yielded speakers to be evaluated as they were uncertain (unsure of their answer). However, we approach this result with caution. Since duration of the stimuli itself might have effected perception of pitch contour. There needs to be comparison with neutral modality of locative pronouns. Moreover, it cannot be claimed that this is a general function of duration. In different context, duration might signal something else. As to the effect of locative pronouns on speaker confidence, the meaning of the lexical item overrode the effect of duration. Perceivers relied on the locative pronouns rather than their durations and inferred from the locative pronouns based on distance-proximity relationship. If we consider the locative pronouns as physical representations of the hidden object according to imaginary scenario, the locative pronoun "burada" (here) which is related to the closest distance caused speakers to be perceived as the most sure by the speakers. This finding is also supported by construal level theory (Trope & Liberman, 2010) and cognitive accessibility (Burenhult, 2003; Kahneman, 2003). As to the effect of epistemic adverbs, the effect of locative

pronouns did not observed. However, we cannot conclude that if a lexical token for certainty/uncertainty is around, it doesn't matter anymore if that token is expressed with certain/uncertain prosody. Since, locative pronouns were requested from speakers to be uttered in neutral modality. Moreover, only determining factor was duration on the perception of speaker confidence degree. In terms of pitch range, in both locative pronouns and locative pronouns which were modified by the epistemic adverbs for certainty, the speakers were perceived as more certain as the pitch range increased. In our data we observed that pitch and stress are related in parallel with the previous studies (Demircan, 1983; Kornfilt, 1997). Furthermore, results showed that the ability of speakers to express their certainty as certain was better than uncertain.

Longer duration is intonational cue for uncertainty whereas shorter duration is for certainty in both perception and production. In the production study, the speakers used high pitch to express certainty and low pitch to express uncertainty for both locative pronouns and epistemic adverbs. However, perception study showed that epistemic adverb "galiba" (probably) was perceived as uncertain with higher pitch. In the production study, speakers preferred higher pitch range to express certainty. Also, when they uttered epistemic adverbs to express the modality (i.e., certain and uncertain) over neutral modality, they preferred high pitch range. In similar, locative pronouns modified by epistemic adverb "kesinlikle" (certainly) perceived as more certain when pitch range was high. This might indicate that pitch level manipulations occurred due to stress on the syllables to be able to convey certainty.

Lastly, our aim was to figure out intonation contours which represent the speaker confidence categories. For the locative pronouns which were uttered by two modalities, we conducted intonational analysis. According to our data, we found some intonation types belong to these speaker confidence categories. The ones mostly occured in our data are as in the following. H* LL% and H*+L LL% intonation contours are the best representers of certainty. Moreover, L* LH% intonation contour is the best representer of uncertainty. The intonation contour H*+L LL% comes in the second order to represent uncertainty. For the in-between category; H*+L LL% intonation contour comes first and L* LL% intonation contour comes in the second order. Boundary tones and duration are the most effective predictors to determine whether the speaker is certain or uncertain. We observed that H* accented tone was used for certainty whereas L* accented tone was used for uncertainty. Moreover, L% boundary tone is certain whereas H% boundary tone is uncertain. We observed that pitch accent affected category of the perceived modality. On the other hand, we observed that LH% boundary tone which was used for questions did not prevent certainty. H* accented tone was found to be more effective than boundary tone. H* LL% and H*+L LL% are not typical contours but they allowed for certain category. L tone decreased certainty.

7.2 Limitations and Future Directions

In our study, we only included female speech since it is easier to track. The female voice gives higher pitch values than male voice. Another reason for this to be easier to track is that female voice is neat due to vowels and stress (Simpson, 2009). Our aim was to compare the frequencies based on one gender. For a future research male voice can be examined as well.

Another limitation in our study was that each speaker utttered the tokens only once. We should have included several trials such as three trials and include the second trial in our experiment as material. Participants would have uttered lexical items three times as in the study of Ipek & Jun (2013) as an answer to the question of the place of the key. Also, the second auditory recording could have been included in our small data corpus. However, we had 7 of the tokens even if they were uttered by different speakers also just having 7 speakers is a limitation. We might have missed some different strategies to express epistemic modality. Also, since these 7 speakers already used different strategies, the results of the listener analysis might reflect this inconsistency or this variability. The fact that most effects turned out insignificant in Experiment 1 might be due to that.

Another limitation is epistemic prosody and adverbs were not tested in naturalistic settings. We had lack of accurate context and our context was only limited to the explanation by the experimenter during elicitation of the utterances whereas for comprehension of the utterances it was screened as an imaginary scenario in the instruction part of the experiment. It is not realistic when compared to real life dialogue, or visual stimuli. Furthermore, while participants were evaluating the short audio clips due to automation they might have forgotten to adhere to scenario explained in the instruction part. For instance, a sentence context might be necessary to elicit ecologically valid data. However, as first step minimalistic production scenario based on single words or phrases may provide a baseline from which future studies could depart. Since our experiments included only one or two word answers, one might want to observe possible outcomes when longer utterances are used. The manipulation of properties such as intonation and delays might become more strain with longer utterances (Dral, Heylen, & op den Akker, 2011). As a future study, another task can be added which is perceptual and thus speakers would be successful to express uncertainty as well.

Finally, we observed that speakers were better at expressing certainty than uncertainty based on the perception of the listeners that rated the items mostly as certain. This might be due to that speakers had an imaginary scenario and created their own cognitive context. As regards self-presentation our speakers more accurately conveyed whether they were sure of the answer rather than unsure. Krahmer & Swerts (2005) in their experiment concluded the reason that children cannot explain uncertainty is due to self-presentation. In our experiments, speakers might be less cared for self-presentation as

well. This is speculative. “Self-presentation” is a rather high-level cognitive term, therefore needs to be treated with caution.

It would be more reliable to conduct a more controlled experiment based on the demonstrative semantics merely by using locative pronouns. Moreover, regarding the fact that the item “burada” was rated as the most certain one, a future study a Demonstrative questionnaire (Wilkins, 1999) would be developed to see usage of these locative pronouns. With respect to duration of an utterance to be able to suggest that value of an utterance is longer or shorter there is need for comparisons between mean and obtained values (Dral, Heylen, & op den Akker, 2011). We can conclude that uncertainty still remains uncertain for us.

REFERENCES

Akaike, H. (1985). Prediction and entropy. In *Selected Papers of Hirotugu Akaike* (pp. 387-410). Springer, New York, NY.

Arnold, J. E., Kaiser, E., Kahn, J. M., & Kim, L. K. (2013). Information structure: linguistic, cognitive, and processing approaches. *Wiley Interdisciplinary Reviews: Cognitive Science*, *4*(4), 403-413.

Audacity Team (2017). Audacity(R): Free Audio Editor and Recorder [Computer application]. Version 2.1.3 retrieved from https://audacityteam.org/

Austin, J. L. (1975). *How to do things with words* (Vol. 88). Oxford university press.

Balog, H. L., & Brentari, D. (2008). The relationship between early gestures and intonation. *First Language*, *28*(2), 141-163.

Barth-Weingarten, D. (2009). *Where prosody meets pragmatics*(Vol. 8). BRILL.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1-9 http://CRAN. *R-project. org/package= lme4*.

Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., & Singmann, H. (2013). Linear mixed-effects models using Eigen and S4. R package version 1.0-5.

Boersma, P. & Weenink, D. (2017). Praat: doing phonetics by computer [Computer program]. Version 6.0.28, retrieved 29 March 2017 from http://www.praat.org/

Bolker, B. M., Brooks, M. E., Clark, C. J., Geange, S. W., Poulsen, J. R., Stevens, M. H. H., & White, J. S. S. (2009). Generalized linear mixed models: a practical guide for ecology and evolution. *Trends in ecology & evolution*, *24*(3), 127-135.

Borràs-Comes, J., del Mar Vanrell, M., & Prieto, P. (2014). The role of pitch range in establishing intonational contrasts. *Journal of the International Phonetic Association*, *44*(1), 1-20.

Borràs-Comes, J., Roseano, P., Vanrell, M. D. M., Chen, A., & Prieto, P. (2011, November). Perceiving uncertainty: facial gestures, intonation, and lexical choice. In *Proceedings of the 2n Conference on Gesture and Speech in Interaction*.

Burenhult, N. (2003). Attention, accessibility, and the addressee. *Pragmatics. Quarterly Publication of the International Pragmatics Association (IPrA)*, *13*(3), 363-379.

Chen, S., Zhu, Y., & Wayland, R. (2017). Effects of stimulus duration and vowel quality in cross-linguistic categorical perception of pitch directions. *PloS one*, *12*(7), e0180656.

Childs, C. (2014). Improve Your American English Accent.

Demircan, Ö. (1983). *Türkçe ezgilemeye giriş*. Türk Tarih Kurumu Basımevi.

Demircan, Ö. (2001). *Türkçenin ezgisi*. Yıldız Teknik Üniversitesi Vakfı.

Demirezen, M. (2014). The tag questions: Certainty versus uncertainty issue in English pitches and intonation in relation to anthropology. *The Anthropologist*, *18*(3), 1059-1067.

Dickey, D. A. (2010, April 11-14). *Ideas and Examples in Generalized Linear Mixed Models*. Paper presented at SAS Global Forum 2010: Statistics and Data Analysis, Seattle, WA.

Dijkstra, C., Krahmer, E., & Swerts, M. (2006). Manipulating uncertainty: The contribution of different audiovisual prosodic cues to the perception of confidence. In *Proceedings of the Speech Prosody Conference*.

Dral, J., Heylen, D., & op den Akker, R. (2011). Detecting uncertainty in spoken dialogues: an exploratory research for the automatic detection of speaker uncertainty by using prosodic markers. In *Affective Computing and Sentiment Analysis* (pp. 67-77). Springer, Dordrecht.

Ergenç, İ. (2002). *Konuşma dili ve türkçenin söyleyiş sözlüğü*. Multilingual.

Faytak, M., & Alan, C. L. (2011). A Typological Study of the Interaction between Level Tones and Duration. In *ICPhS* (pp. 659-662).

Fernald, A. (1989). Intonation and communicative intent in mothers' speech to infants: Is the melody the message?. *Child development*, 1497-1510.

Fernald, A., & Simon, T. (1984). Expanded intonation contours in mothers' speech to newborns. *Developmental psychology*, 20(1), 104.

Friend, M. (2001). The transition from affective to linguistic meaning. *First Language, 21*(63), 219-243.

Friend, M. (2003). What should I do? Behavior regulation by language and paralanguage in early childhood. *Journal of Cognition and Development*, *4*(2), 161-183.

Gandour, J. (1977). On the interaction between tone and vowel length: Evidence from Thai dialects. *Phonetica*, *34*(1), 54-65.

Gelman, A., & Hill, J. (2006). *Data analysis using regression and multilevel/hierarchical models*. Cambridge university press.

Gravano, A., Benus, S., Hirschberg, J., German, E. S., & Ward, G. (2008, May). The effect of contour type and epistemic modality on the assessment of speaker certainty. In *proc. international conference: Speech prosody*.

Green, M. G. (1979). The developmental relation between cognitive stage and the comprehension of speaker uncertainty. *Child Development*, 666-674.

Green, M. S., & Williams, J. N. (Eds.). (2007). *Moore's paradox: new essays on belief, rationality, and the first person*. Oxford University Press on Demand.

Grice, H. P. (1968). Utterer's meaning, sentence-meaning, and word-meaning. In *Philosophy, Language, and Artificial Intelligence* (pp. 49-66). Springer, Dordrecht.

Grice, M., & Savinot, M. (1997). Can pitch accent type convey information status in yes-no questions?. *Concept to Speech Generation Systems*.

Gumperz, J. J., & Levinson, S. C. (1991). Rethinking linguistic relativity. *Current Anthropology*, *32*(5), 613-623.

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

Göksel, A., Kelepir, M., & Üntak-Tarhan, A. (2007). Türkçe soru cümlelerinde ezgi. *Y. Aksan, and M. Aksan (Eds.), 21, 309-316.*

Göksel, A., & Kerslake, C. (2004). *Turkish: A comprehensive grammar*. Routledge.

Göksel, A., & Özsoy, A. S. (2000). Is there a focus position in Turkish. *Studies on Turkish and Turkic languages*, 219-228.

Hadding-Koch, K., & Studdert-Kennedy, M. (1964). An experimental study of some intonation contours. *Phonetica*, *11*(3-4), 175-185.

Hart, J. T. (1965). Memory and the feeling-of-knowing experience. *Journal of educational psychology*, *56*(4), 208.

Hirschberg, J. (2004). Pragmatics and intonation. *The handbook of pragmatics*, 515-537.

Hirschberg, J., & Pierrehumbert, J. (1986, July). The intonational structuring of discourse. In *Proceedings of the 24th annual meeting on Association for Computational Linguistics* (pp. 136-144). Association for Computational Linguistics.

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

Hirst, D., & Di Cristo, A. (Eds.). (1998). *Intonation systems: a survey of twenty languages*. Cambridge University Press.

Holmes, J. (1982). Expressing doubt and certainty in English. *RELC journal*, *13*(2), 9-28.

Hübscher, I., Esteve-Gibert, N., Igualada, A., & Prieto, P. (2017). Intonation and gesture as bootstrapping devices in speaker uncertainty. *First Language*, *37*(1), 24-41.

Ipek, C., & Jun, S. A. (2013, June). Towards a model of intonational phonology of Turkish: Neutral intonation. In *Proceedings of Meetings on Acoustics ICA2013* (Vol. 19, No. 1, p. 060230). ASA.

Ipek, C., & Jun, S. A. (2014). Distinguishing phrase-final and phrase-medial high tone on finally stressed words in Turkish. In *The Proceedings of the 7th Speech Prosody International Conference. Dublin: Ireland.[http://www. linguistics. ucla. edu/people/jun/papers% 20in% 20pdf/Ipe k_Jun-SpeechProsody-2014-Turkish. pdf, accessed 01/15]*.

Janssen, T. A. J. M. (1995). Deixis from a cognitive point of view. *TRENDS IN LINGUISTICS STUDIES AND MONOGRAPHS*, *84*, 245-270.

Jiang, X., & Pell, M. D. (2017). The sound of confidence and doubt. *Speech Communication*, *88*, 106-126.

Kahneman, D. (2003). A perspective on judgment and choice: mapping bounded rationality. *American psychologist*, *58*(9), 697.

Kan, S. (2009). Prosodic Domains and the syntax-prosody mapping in Turkish. *MA diss. Boğaziçi University.*

Kamalı, B. (2011). Topics at the PF interface of Turkish. *Unpublished PhD thesis, Harvard University.*

Kornfilt, J. (1997). *Turkish*. London: Routledge.

Krahmer, E., & Swerts, M. (2005). How children and adults produce and perceive uncertainty in audiovisual speech. *Language and speech, 48*(1), 29-53.

Küntay, A. C., & Özyürek, A. (2006). Learning to use demonstratives in conversation: what do language specific strategies in Turkish reveal?. *Journal of Child Language, 33*(2), 303-320.

Ladd, D. R. (1996). *Intonational phonology*. Cambridge: Cambridge University Press.

Leech, G., & Svartvik, J. (1975). *A communicative grammar of English*. London.

Lewis, G. L. (1970). Turkish grammar.

Martha C. P. 1996. *Phonology In English Language Teaching.* London: Longman

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior research methods, 44*(2), 314-324.

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing Type I error and power in linear mixed models. *Journal of Memory and Language, 94*, 305-315.

McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. University of Chicago press.

Meira, S. (2003). 'Addressee effects' in demonstrative systems. *Deictic Conceptualisation of Space, Time and Person, 112*, 3.

Moore, C., Bryant, D., & Furrow, D. (1989). Mental terms and the development of certainty. *Child Development*, 167-171.

Moore, C., Harris, L., & Patriquin, M. (1993). Lexical and prosodic eues in the comprehension of relative certainty. *Journal of Child Language, 20*(1), 153-167.

Moore, C., Pure, K., & Furrow, D. (1990). Children's understanding of the modal expression of speaker certainty and uncertainty and its relation to the development of a representational theory of mind. *Child development, 61*(3), 722-730.

Nolan, F. (2014). Intonation.

Özge, U. (2003). A tune-based account of Turkish information structure. *Ankara: Middle East Technical University, Ankara master's thesis. Online: http://www. ii. metu. edu. tr/~ umut*.

Özge, U., & Bozsahin, C. (2010). Intonation in the grammar of Turkish. *Lingua*, *120*(1), 132-175.

Özyurek, A. (1998). An analysis of the basic meaning of Turkish demonstratives in face-to-face conversational interaction. In *Oralité et gestualité: Communication multimodale, interteraction*(pp. 609-614). L'Harmattan.

Palmer, F. R. (2001). *Mood and modality*. Cambridge University Press.

Palmer, F. R. (2014). *Modality and the English modals*. Routledge.

Pickering, M. J., & Garrod, S. (2007). Do people use language production to make predictions during comprehension?. *Trends in cognitive sciences*, *11*(3), 105-110.

Pierrehumbert, J. B. (1980). *The phonology and phonetics of English intonation* (Doctoral dissertation, Massachusetts Institute of Technology).

Pierrehumbert, J., & Hirschberg, J. B. (1990). The meaning of intonational contours in the interpretation of discourse. *Intentions in communication*, 271-311.

Pinheiro, J. C., & Bates, D. M. (2000). Mixed-Effects Models in S and S PLUS. New York: Springer.

Piwek, P., Beun, R. J., & Cremers, A. (2008). 'Proximal'and'distal'in language and cognition: Evidence from deictic demonstratives in Dutch. *Journal of Pragmatics*, *40*(4), 694-718.

R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Roseano, P., González, M., Borràs-Comes, J., & Prieto, P. (2016). Communicating epistemic stance: How speech and gesture patterns reflect epistemicity and evidentiality. *Discourse Processes*, *53*(3), 135-174.

Savino, M., & Grice, M. (2007, August). The role of pitch range in realising pragmatic contrasts–The case of two question types in Italian. In *Proceedings of the XVIth International Congress of Phonetic Sciences* (pp. 1037-1040). Saarbrücken, Germany: Pirrot GmbH.

Searle, J. R. (1976). A classification of illocutionary acts. *Language in society*, *5*(1), 1-23.

Simpson, A. P. (2009). Phonetic differences between male and female speech. *Language and Linguistics Compass*, 3(2), 621-640.

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

Swerts, M., & Krahmer, E. (2005). Audiovisual prosody and feeling of knowing. *Journal of Memory and Language*, *53*(1), 81-94.

Trope, Y., & Liberman, N. (2010). Construal-level theory of psychological distance. *Psychological review*, *117*(2), 440.

Underhill, R. (1976). *Turkish grammar* (p. xviii474). Cambridge, MA: Mit Press.

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

Wilkins, D. (1999). The 1999 demonstrative questionnaire:"THIS" and "THAT" in comparative perspective. In *Manual for the 1999 field season* (pp. 1-24). Max Planck Institute for Psycholinguistics.

Winter, B. (2013). Linear models and linear mixed effects models in R with linguistic applications. *arXiv preprint arXiv:1308.5499*.

Yu, A. C. (2010). Tonal effects on perceived vowel duration. *Laboratory phonology*, *10*, 151-168.

Yu, A. C., Lee, H., & Lee, J. (2014). Variability in perceived duration: pitch dynamics and vowel quality. *In Fourth International Symposium on Tonal Aspects of Languages.*

APPENDICES

APPENDIX A

METU Ethics Committee Approval

UYGULAMALI ETİK ARAŞTIRMA MERKEZİ
APPLIED ETHICS RESEARCH CENTER

ORTA DOĞU TEKNİK ÜNİVERSİTESİ
MIDDLE EAST TECHNICAL UNIVERSITY

DUMLUPINAR BULVARI 06800
ÇANKAYA ANKARA/TURKEY
T: +90 312 210 22 91
F: +90 312 210 79 59
ueam@metu.edu.tr
www.ueam.metu.edu.tr

Sayı: 28620816 /353

06 Haziran 2018

Konu: Değerlendirme Sonucu

Gönderen: ODTÜ İnsan Araştırmaları Etik Kurulu (İAEK)

İlgi: İnsan Araştırmaları Etik Kurulu Başvurusu

Sayın Doç.Dr. Annette HOHENBERGER ve Dr. Umut ÖZGE

Danışmanlığını yaptığınız yüksek lisans öğrencisi Ayşenur HÜLAGÜ'nün **"Production and perception of intonational cues expressing epistemic confidence in Turkish"** başlıklı araştırması İnsan Araştırmaları Etik Kurulu tarafından uygun görülerek gerekli onay **2018-FEN-002** protokol numarası ile **08.06.2018 - 30.12.2018** tarihleri arasında geçerli olmak üzere verilmiştir.

Bilgilerinize saygılarımla sunarım.

Prof. Dr. Ş. Halil TURAN
Başkan V

Prof. Dr. Ayhan SOL
Üye

Prof. Dr. Ayhan Gürbüz DEMİR
Üye

Doç. Dr. Yaşar KONDAKÇI
Üye

Doç. Dr. Zana ÇITAK
Üye

Doç. Dr. Emre SELÇUK
Üye

Dr. Öğr. Üyesi Pınar KAYGAN
Üye

APPENDIX B

Consent Form

Onam Formu

Bu araştırma Orta Doğu Teknik Üniversitesi Bilişsel Bilimler Ana Bilim Dalı Yüksek Lisans Programı COGS 599 kodlu 'Yüksek Lisans Tezi' dersi kapsamında Ayşenur Hülagü tarafından, Doç. Dr. Annette Hohenberger danışmanlığında yürütülmektedir. Araştırmanın amacı kişilerin kesinlik ve belirsizlik durumlarını hangi yollarla ifade ettiklerini saptamaktır. Bu amaçla kesinlik ve belirsizlik durumlarında ortaya koyduğumuz ezgi örüntüleri saptanıp sınıflandırılacaktır. Bu gerçekleştirilirken olasılık zarfları ile işaret zamirlerinden faydalanılacaktır. Araştırma, sunulacak senaryoda zarf ve zamirlerin seslendirilmesi ve bu esnadaki ses kaydından oluşmaktadır. Elde edilen ses kayıtları deneyin ikinci çalışmasında konuşmacıların dinleyiciler tarafından nasıl algılandığını saptamak üzere kullanılacaktır.

Bizimle paylaştığınız bilgileri ve araştırmamızda kullanmamız için vermiş olduğunuz onayı aşağıda belirtilen mail adresine bildirerek her an iptal ettirebilirsiniz. Onayınızı iptal ederseniz araştırmacılar sizin paylaştığınız bilgileri kullanmayacaklardır. Bu araştırmayla ilgili daha fazla bilgi edinmek isterseniz +90 536 920 86 39 numaralı telefonu arayabilir ya da aysenur.hulagu@gmail.com adresine e-posta gönderebilirsiniz.

Bu araştırmaya tamamen gönüllü olarak katılıyorum ve istediğim zaman yarıda kesip çıkabileceğimi biliyorum. Vereceğim bilgilerin kimliğimle eşleştirilmeyeceğini biliyor ve bilimsel amaçlı yayımlarda kullanılmasını kabul ediyorum.

İsim – Soyad:

Tarih: __/__/__

İmza:

APPENDIX C

Instruction

Yönerge

Size sunulan farazi senaryoda anahtar odada bir yere gizlenmiş durumdadır. Odaya giren araştırmacı size anahtarın yerini sorar.

a) Anahtarın yerini biliyorsunuz ve nerede olduğundan eminsiniz.
b) Anahtarın yerini bilmiyorsunuz ve nerede olduğundan emin değilsiniz.

İlk durumda anahtarın nerede olduğunu biliyorsunuz ve eminsiniz. Odaya giren kişi size anahtarın nerede olduğunu soruyor. Odaya giren kişiye yalnızca "burada", "şurada", ve "orada" işaret zamirlerini kullanarak cevap vermeniz gerekiyor.

İkinci durumda anahtarın nerede olduğunu bilmiyorsunuz ve emin değilsiniz. Odaya giren kişi size anahtarın nerede olduğunu soruyor. Odaya giren kişiye yalnızca "burada", "şurada", ve "orada" işaret zamirlerini kullanarak cevap vermeniz gerekiyor.

Bunun yanı sıra "kesinlikle" olasılık zarfı ile cevap vereceksiniz.

Bunun yanı sıra "galiba" olasılık zarfı ile cevap vereceksiniz.

Sonraki kısımda ise "kesinlikle" olasılık zarfı ile sırayla "burada", "şurada", ve "orada" işaret zamirlerini kombinleyeceksiniz. Yani "kesinlikle burada", "kesinlikle şurada", ve "kesinlikle orada" şeklinde olacak.

Daha sonraki kısımda ise "galiba" olasılık zarfı ile sırayla "burada", "şurada", ve "orada" işaret zamirlerini kombinleyeceksiniz. Yani "galiba burada", "galiba şurada", ve "galiba orada" şeklinde olacak.

1) Anahtar bir yere gizlenmiş durumda. Ben odaya giriyorum ve anahtarın nerede olduğunu soruyorum. Anahtarın yerini biliyorsunuz ve nerede olduğundan eminsiniz. Emin olduğunuz bir şekilde "burada" sözcüğü ile cevap verebilir misiniz?

2) Anahtar bir yere gizlenmiş durumda. Ben odaya giriyorum ve anahtarın nerede olduğunu soruyorum. Anahtarın yerini biliyorsunuz ve nerede olduğundan eminsiniz. Emin olduğunuz bir şekilde "orada" sözcüğü ile cevap verebilir misiniz?

3) Anahtar bir yere gizlenmiş durumda. Ben odaya giriyorum ve anahtarın nerede olduğunu soruyorum. Anahtarın yerini biliyorsunuz ve nerede olduğundan eminsiniz. Emin olduğunuz bir şekilde "şurada" sözcüğü ile cevap verebilir misiniz?

4) Anahtar bir yere gizlenmiş durumda. Ben odaya giriyorum ve anahtarın nerede olduğunu soruyorum. Siz anahtarın yerini bilmiyorsunuz ve nerede olduğundan emin değilsiniz. Emin olmadığınız bir şekilde "burada" sözcüğü ile cevap verebilir misiniz?

5) Anahtar bir yere gizlenmiş durumda. Ben odaya giriyorum ve anahtarın nerede olduğunu soruyorum. Siz anahtarın yerini bilmiyorsunuz ve nerede olduğundan emin değilsiniz. Emin olmadığınız bir şekilde "orada" sözcüğü ile cevap verebilir misiniz?

6) Anahtar bir yere gizlenmiş durumda. Ben odaya giriyorum ve anahtarın nerede olduğunu soruyorum. Siz anahtarın yerini bilmiyorsunuz ve nerede olduğundan emin değilsiniz. Emin olmadığınız bir şekilde "şurada" sözcüğü ile cevap verebilir misiniz?

7) "Kesinlikle" sözcüğünü emin olduğumuz durumlarda kullanırız. Emin olduğunuz bir şekilde "kesinlikle" diyebilir misiniz ?

8) "Galiba" sözcüğünü emin olmadığımız durumlarda kullanırız. Emin olmadığınız bir şekilde "galiba" diyebilir misiniz ?

9) "Galiba" sözcüğünü nötr bir şekilde söyleyebilir misiniz ?

10) "Kesinlikle" sözcüğünü nötr bir şekilde söyleyebilir misiniz ?

11) Anahtar bir yere gizlenmiş durumda. Ben odaya giriyorum ve anahtarın nerede olduğunu soruyorum. Anahtarın yerini biliyorsunuz ve nerede olduğundan eminsiniz. "Kesinlikle" sözcüğü ile "burada" sözcüğünü birleştireceğiz. "kesinlikle" sözcüğünde emin olduğunuzun vurgusu "burada" sözcüğü ise nötr bir şekilde cevap verebilir misiniz?

12) Anahtar bir yere gizlenmiş durumda. Ben odaya giriyorum ve anahtarın nerede olduğunu soruyorum. Anahtarın yerini biliyorsunuz ve nerede olduğundan eminsiniz. "Kesinlikle" sözcüğü ile "orada" sözcüğünü birleştireceğiz.

“kesinlikle” sözcüğünde emin olduğunuzun vurgusu “orada” sözcüğü ise nötr bir şekilde cevap verebilir misiniz?

13) Anahtar bir yere gizlenmiş durumda. Ben odaya giriyorum ve anahtarın nerede olduğunu soruyorum. Anahtarın yerini biliyorsunuz ve nerede olduğundan eminsiniz. “Kesinlikle” sözcüğü ile “şurada” sözcüğünü birleştireceğiz. “kesinlikle” sözcüğünde emin olduğunuzun vurgusu “şurada” sözcüğü ise nötr bir şekilde cevap verebilir misiniz?

14) Anahtar bir yere gizlenmiş durumda. Ben odaya giriyorum ve anahtarın nerede olduğunu soruyorum. Siz anahtarın yerini bilmiyorsunuz ve nerede olduğundan emin değilsiniz. “Galiba” sözcüğü ile “burada” sözcüğünü birleştireceğiz. “galiba” sözcüğünde emin olmadığınızın vurgusu “burada” sözcüğü ise nötr bir şekilde cevap verebilir misiniz?

15) Anahtar bir yere gizlenmiş durumda. Ben odaya giriyorum ve anahtarın nerede olduğunu soruyorum. Siz anahtarın yerini bilmiyorsunuz ve nerede olduğundan emin değilsiniz. “Galiba” sözcüğü ile “şurada” sözcüğünü birleştireceğiz. “galiba” sözcüğünü emin olmadığınızın vurgusu “şurada” sözcüğü ise nötr bir şekilde cevap verebilir misiniz?

16) Anahtar bir yere gizlenmiş durumda. Ben odaya giriyorum ve anahtarın nerede olduğunu soruyorum. Siz anahtarın yerini bilmiyorsunuz ve nerede olduğundan emin değilsiniz. “Galiba” sözcüğü ile “orada” sözcüğünü birleştireceğiz. “galiba” sözcüğünde emin olmadığınızın vurgusu “orada” sözcüğü ise nötr bir şekilde cevap verebilir misiniz?

APPENDIX D

Debriefing Form

Bilgilendirme Formu

Bu araştırma daha önce de belirtildiği gibi Orta Doğu Teknik Üniversitesi Bilişsel Bilimler Ana Bilim Dalı Yüksek Lisans Programı COGS 599 kodlu 'Yüksek Lisans Tezi' dersi kapsamında Ayşenur Hülagü tarafından, Doç. Dr. Annette Hohenberger danışmanlığında yürütülmektedir. Çalışmamızda epistemik yani bilgi eksikliğinin yarattığı bir belirsizlik durumunda iletişimdeki bireylerin ezgisel ipuçlarından hangi ya da hangilerine güvenerek bu belirsizliği çözmeye çalıştıklarını bulmayı amaçlıyoruz.

Bu nedenle bilgi içerme işlevi olan ezgiyi epistemik bağlamda incelemeyi seçtik. Size sunduğumuz senaryoda amacımız işaret zamirlerini ("burada", "şurada", ve "orada") farklı ezgilerle kombinleyerek kesinlik ve belirsizlik durumları yaratmaktır. Ek olarak sözcüksel düzeyde kesinlik ve belirsizlik ifade edebilmek için de olasılık zarflarına ("kesinlikle" ve "galiba") yer verdik. Çalışmanın sonunda elde edilen ses kayıtlarını inceleyerek kesinlik ve belirsizlik durumları konuşma dilinde ifade edilirken kullanılan ezgi örüntülerini saptamak istiyoruz.

Bu çalışmadan alınacak ilk verilerin Mart 2018 sonunda elde edilmesi amaçlanmaktadır. Elde edilen bilgiler sadece bilimsel araştırma ve yazılarda kullanılacaktır. Bu araştırmaya katıldığınız için tekrar çok teşekkür ederiz.

Araştırmanın sonuçlarını öğrenmek ya da daha fazla bilgi almak için aşağıdaki isimlere başvurabilirsiniz.

Araştrmacı: Ayşenur Hülagü

e-posta: aysenur.hulagu@gmail.com

Çalışmaya katkıda bulunan bir gönüllü olarak katılımcı haklarınızla ilgili veya etik ilkelerle ilgi soru veya görüşlerinizi ODTÜ Uygulamalı Etik Araştırma Merkezi'ne iletebilirsiniz.

e-posta: ueam@metu.edu.tr

APPENDIX E

Consent Form

Onam Formu

Bu araştırma Orta Doğu Teknik Üniversitesi Bilişsel Bilimler Ana Bilim Dalı Yüksek Lisans Programı COGS 599 kodlu ‘Yüksek Lisans Tezi' dersi kapsamında Ayşenur Hülagü tarafından, Doç. Dr. Annette Hohenberger danışmanlığında yürütülmektedir. Araştırmanın amacı kişilerin kesinlik ve belirsizlik durumlarını hangi yollarla ifade ettiklerini saptamaktır. Araştırmada size sunulacak olan ses kliplerindeki materyaller günlük yaşamınızda karşılaşmakta olduğunuz kelimelerden oluşmaktadır. Araştırma süresince kulaklık kullanmanızı rica ediyoruz.

Bizimle paylaştığınız bilgileri araştırmamızda kullanmamız için vermiş olduğunuz onayı aşağıda belirtilen mail adresine bildirerek her an iptal ettirebilirsiniz. Onayınızı iptal ederseniz araştırmacılar sizin paylaştığınız bilgileri kullanmayacaklardır. Bu araştırmayla ilgili daha fazla bilgi edinmek isterseniz +90 536 920 86 39 numaralı telefonu arayabilir ya da aysenur.hulagu@gmail.com adresine e-posta gönderebilirsiniz.

Bu araştırmaya tamamen gönüllü olarak katılıyorum ve istediğim zaman yarıda kesip çıkabileceğimi biliyorum. Vereceğim bilgilerin kimliğimle eşleştirilmeyeceğini biliyor ve bilimsel amaçlı yayımlarda kullanılmasını kabul ediyorum.

İsim – Soyad:

Tarih: __/__/__

İmza:

APPENDIX F

Debriefing Form

Bilgilendirme Formu

Bu araştırma daha önce de belirtildiği gibi Orta Doğu Teknik Üniversitesi Bilişsel Bilimler Ana Bilim Dalı Yüksek Lisans Programı COGS 599 kodlu 'Yüksek Lisans Tezi' dersi kapsamında Ayşenur Hülagü tarafından, Doç. Dr. Annette Hohenberger danışmanlığında yürütülmektedir. Çalışmamızda epistemik yani bilgi eksikliğinin yarattığı bir belirsizlik durumunda iletişimdeki bireylerin ezgisel ipuçlarından hangi ya da hangilerine güvenerek bu belirsizliği çözmeye çalıştıklarını bulmayı amaçlıyoruz.

Bu nedenle bilgi içerme işlevi olan ezgiyi epistemik bağlamda incelemeyi seçtik. Size sunduğumuz senaryoda amacımız işaret zamirlerini ("burada", "şurada", ve "orada") farklı ezgilerle kombinleyerek kesinlik ve belirsizlik durumları yaratmaktır. Ek olarak sözcüksel düzeyde kesinlik ve belirsizlik ifade edebilmek için de olasılık zarflarına ("kesinlikle" ve "galiba") yer verdik. Çalışmanın sonunda elde edilen yanıtlar değerlendirilerek kesinlik ve belirsizlik ifade ederken konuşmacıların kullandıkları farklı stratejilerden hangisi ya da hangilerinin dinleyiciler tarafından en çok tercih edildiği belirlenecektir.

Bu çalışmadan alınacak ilk verilerin Mart 2018 sonunda elde edilmesi amaçlanmaktadır. Elde edilen bilgiler sadece bilimsel araştırma ve yazılarda kullanılacaktır. Bu araştırmaya katıldığınız için tekrar çok teşekkür ederiz.

Araştırmanın sonuçlarını öğrenmek ya da daha fazla bilgi almak için aşağıdaki isimlere başvurabilirsiniz.

Araştrmacı: Ayşenur Hülagü

e-posta: aysenur.hulagu@gmail.com

Çalışmaya katkıda bulunan bir gönüllü olarak katılımcı haklarınızla ilgili veya etik ilkelerle ilgi soru veya görüşlerinizi ODTÜ Uygulamalı Etik Araştırma Merkezi'ne iletebilirsiniz.

e-posta: ueam@metu.edu.tr

APPENDIX G

Model Development

Experiment 2 (Perception Study)

Block I

Speaker Confidence (Perceived Modality)

The intercept only model was created by adding random intercepts for participants and items to predict speaker confidence, controlling for by-participant and by-item variability.

Intercept only model is as the following:

Speaker_Confidence ~ 1 + (1 | Participant.no) + (1 | item)

The model started to be constructed with the fixed structure. The effects of the acoustic properties on speaker confidence (dependent variable) was analysed in the direction of the following question: Whether acoustic properties were predictors of speaker confidence?

A model was created that included the fixed effect "CMPitch" (centered mean pitch in Hertz) to predict speaker confidence, controlling for by-participant and by-item variability.

The model is as the following:

Speaker_Confidence ~ CMPitch + (1 | Participant.no) + (1 | item)

The two models, the null and the one with the factor "CMPitch", were compared by performing the likelihood ratio test. The anova ( ) function did not yield a statistically significant difference between these two models. Mean pitch was found ineffective on speaker confidence ($\chi^2(1) = 0.0028$, $p = 0.9578$) and was not included in the model.

Another model was created that included the fixed effect "C.Dur" (centered duration in milliseconds) to predict speaker confidence, controlling for by-participant and by-item variability.

The model is as the following:

Speaker_Confidence ~ C.Dur + (1 | Participant.no) + (1 | item)

The two models, the null and the one with the factor "C.Dur", were compared by performing the likelihood ratio test. The anova ( ) function yielded a statistically significant difference between these two models. Duration was found effective on speaker confidence ($\chi^2(1) = 13.185$, $p = 0.0002822$) and kept in the model.

Another model was created that included the fixed effect "CPRange" (centered pitch range in Hertz) to predict speaker confidence, controlling for by-participant and by-item variability.

The model is as the following:

Speaker_Confidence ~ CPRange + (1 | Participant.no) + (1 | item)

The two models, the null and the one with the factor "CPRange", were compared by performing the likelihood ratio test. The anova ( ) function did not yield a statistically significant difference between these two models. Pitch range was found ineffective on speaker confidence ($\chi^2(1) = 0.0641$, $p = 0.8001$) and was not included in the model.

Moreover, the models, the one with the factor "C.Dur" and the one with the factors "C.Dur" and "CMPitch" were compared. In this comparison, duration was controlling factor while mean pitch was tested. Mean pitch did not contribute to the model included only duration ($\chi^2(1) = 0.4973$, $p = 0.4807$). Furthermore, the models, the one with the factor "C.Dur" and the one with the factors "C.Dur" and "CPRange" were compared. In this comparison, duration was controlling factor while pitch range was tested. Pitch range contributed to the model included only duration ($\chi^2(1) = 6.5043$, $p = 0.01076$). Lastly, the models, the one with the factors "C.Dur" and "CPRange" and the one with the factors "C.Dur", "CPRange", and "CMPitch" were compared. In this comparison, duration and pitch range were controlling factors while mean pitch was tested. Mean pitch did not contribute to the model included duration and pitch range ($\chi^2(1) = 0.7572$, $p = 0.3842$). According to the results of the comparison of the models, among acoustic properties duration and pitch range were found as predictor of speaker confidence and kept in the model.

The model is as the following:

Speaker_Confidence ~ C.Dur + CPRange (1 | Participant.no) + (1 | item)

The model continued to be constructed with the fixed structure. The effect of the locative pronouns on speaker confidence (dependent variable) was analysed in the direction of the following question: Whether locative pronoun was the predictor of speaker confidence?

A model was created that included the fixed effect "Locative" to predict speaker confidence, controlling for by-participant and by-item variability.

The model is as the following:

Speaker_Confidence ~ Locative + (1 | Participant.no) + (1 | item)

The two models, the null and the one with the factor "Locative", were compared by performing the likelihood ratio test. The anova ( ) function did not yield a statistically significant difference between these two models. Locative was found ineffective on speaker confidence ($\chi^2(2) = 0.9803$, $p = 0.6125$) and was not kept in the model.

In addition, the models, the one with the factor "C.Dur" and the one with the factors "C.Dur" and "Locative" were compared. In this comparison, duration was controlling factor while locative pronouns were tested. Locative contributed to the model included only duration ($\chi^2(2) = 6.546$, $p = 0.03789$). Accordingly, an alternative model was constructed.

The model is as the following:

Speaker_Confidence ~ C.Dur + Locative (1 | Participant.no) + (1 | item)

A model was created that included the fixed effect "Locative" in addition to the fixed effects "C.Dur" and "CPRange" to predict speaker confidence, controlling for by-participant and by-item variability.

The model is as the following:

Speaker_Confidence ~ C.Dur + CPRange + Locative (1 | Participant.no) + (1 | item)

The models, the one with the factors "C.Dur" and "CPRange" and the one with the factors "C.Dur", "CPRange", and "Locative" were compared. In this comparison, duration and pitch range were controlling factors while locative pronoun was tested. The model was nearly unidentifiable. Hence, locative did not take part in the model as one of the predictors.

The model continued to be constructed with the fixed structure. The effect of modality on speaker confidence (dependent variable) was analysed in the direction of the following question: Whether modality was the predictor of speaker confidence?

A model was created that included the fixed effect “Modality” in addition to the fixed effects “C.Dur” and “CPRange” to predict speaker confidence, controlling for by-participant and by-item variability.

The model is as the following:

Speaker_Confidence ~ C.Dur + CPRange + Modality (1|Participant.no) + (1|item)

The models, the one with the factors “C.Dur” and “CPRange” and the one with the factors “C.Dur”, “CPRange”, and “Modality” were compared. In this comparison, duration and pitch range were controlling factors while modality was tested. The model was nearly unidentifiable. Hence, modality did not take part in the model as one of the predictors.

Another model was created that included the fixed effect “Modality” in addition to the fixed effects “C.Dur” and “Locative” to predict speaker confidence, controlling for by-participant and by-item variability.

The model is as the following:

Speaker_Confidence ~ C.Dur + Locative + Modality (1|Participant.no) + (1|item)

The models, the one with the factors “C.Dur” and “Locative” and the one with the factors “C.Dur”, “Locative”, and “Modality” were compared. In this comparison, duration and locative were controlling factors while modality was tested. The model failed to converge. Hence, modality did not take part in the model as one of the predictors.

As a last step, to finalize construction of the model with the fixed structure, interaction between the fixed factors was controlled. Interaction model with the interaction between the factors “C.Dur” and “CPRange” failed to converge. Also, the model with the interaction between the factors “C.Dur” and “Locative” was nearly inidentifiable. Hence, the interaction did not take part in neither of the models.

Selected fixed structure models are as the following:

Speaker_Confidence ~ C.Dur + CPRange + (1 | Participant.no) + (1 | item)

Speaker_Confidence ~ C.Dur + Locative + (1 | Participant.no) + (1 | item)

Since the construction of the model with the fixed structure was finalized, it continued with the random structure. The model that was shown above is the random intercept model. However, the effects of duration, pitch range, and locative are not the same for all participants. Accordingly, the variables on the fixed structure were inserted into participant.no slope yet for the items these covariates (independent variables) were not

changing. Therefore, item was taken as intercept only. Random structure included the fixed effect "C.Dur" and "CPRange" as random slope inserted into participant.no, respectively (i.e., (0 + C.Dur | Participant.no) and (0 + CPRange | Participant.no)). The results did not yield statistically significant differences between random intercept and random slope models, ($\chi^2(1) = 0$, $p = 1$). Random structure included the fixed effect "C.Dur" and "Locative" as random slope inserted into participant.no, respectively (i.e., (0 + C.Dur | Participant.no) and (0 + Locative | Participant.no)). The results did not yield statistically significant differences between random intercept and random slope models ($\chi^2(1) = 0$, $p = 1$ and $\chi^2(6) = 5.7655$, $p = 0.45$).

Note that AIC (Akaike Information Criterion) is used when two models are not nested. The model selection should be based on low value of AIC. This value should decrease when a factor added to the model. Thus, it can be claimed that when AIC is low, the model is better (Akaike, 1985).

At the end of the fixed and random structuring the selected models are as the following:

Formula 0: Speaker_Confidence ~ C.Dur + (1 | Participant.no) + (1 | item)

AIC = 1489.8

Formula 1: Speaker_Confidence ~ C.Dur + CPRange + (1 | Participant.no) + (1 | item)

AIC = 1485.3

Formula 2: Speaker_Confidence ~ C.Dur + Locative + (1 | Participant.no) + (1 | item)

AIC = 1487.3

Correctness (Accuracy)

The intercept only model was created by adding random intercepts for participants and items to predict correctness, controlling for by-participant and by-item variability.

Intercept only model is as the following:

Correctness ~ 1 + (1 | Participant.no) + (1 | item)

The model started to be constructed with the fixed structure. The effects of the acoustic properties on correctness (dependent variable) were analysed in the direction of the following question: Whether acoustic properties were predictors of correctness?

A model was created that included the fixed effect "CMPitch" (centered mean pitch in Hertz) to predict correctness, controlling for by-participant and by-item variability.

The model is as the following:

Correctness ~ CMPitch + (1 | Participant.no) + (1 | item)

The two models, the null and the one with the factor "CMPitch", were compared performing the likelihood ratio test. The anova ( ) function yielded a statistically significant difference between these two models. Mean pitch was found effective on correctness ($\chi^2(1) = 5.0045$, $p = 0.02528$) and kept in the model.

Another model was created that included the fixed effect "C.Dur" (centered duration in milliseconds) to predict correctness, controlling for by-participant and by-item variability.

The model is as the following:

Correctness ~ C.Dur + (1 | Participant.no) + (1 | item)

The two models, the null and the one with the factor "C.Dur", were compared performing the likelihood ratio test. The anova ( ) function did not yield a statistically significant difference between these two models. Duration was found ineffective on correctness ($\chi^2(1) = 0.1839$, $p = 0.668$) and was not included in the model.

Another model was created that included the fixed effect "CPRange" (centered pitch range in Hertz) to predict correctness, controlling for by-participant and by-item variability.

The model is as the following:

Correctness ~ CPRange + (1 | Participant.no) + (1 | item)

The two models, the null and the one with the factor "CPRange", were compared performing the likelihood ratio test. The anova ( ) function did not yield a statistically significant difference between these two models. Pitch range was found ineffective on correctness ($\chi^2(1) = 0.0667$, $p = 0.7962$) and was not included in the model.

Moreover, the models, the one with the factor "CMPitch" and the one with the factors "CMPitch" and "C.Dur" were compared. In this comparison, mean pitch was controlling factor while duration was tested. Duration did not contribute to the model included only mean pitch ($\chi^2(1) = 0.8608$, $p = 0.3535$). Moreover, the models, the one with the factor "CMPitch" and the one with the factors "CMPitch" and "CPRange" were compared. In this comparison, mean pitch was controlling factor while pitch range was tested. Pitch range did not contribute to the model included only mean pitch ($\chi^2(1) = 0.2031$, $p = 0.6522$). Furthermore, the models, the one with the factor "CMPitch" and the one with the factors "CMPitch", "C.Dur", and "CPRange" were compared. In this comparison, mean pitch was controlling factor while duration and pitch range were tested. Duration

and pitch range did not contribute to the model included only mean pitch ($\chi^2(2) = 0.8735$, $p = 0.6461$). According to the results of the comparison of the models, among acoustic properties mean pitch was found as predictor of correctness and kept in the model.

The model is as the following:

Correctness ~ CMPitch + (1 | Participant.no) + (1 | item)

The model continued to be constructed with the fixed structure. The effect of the locative pronouns on correctness (dependent variable) was analysed in the direction of the following question: Whether locative was the predictor of correctness?

A model was created that included the fixed effect "Locative" in addition to the fixed effect "CMPitch" to predict correctness, controlling for by-participant and by-item variability.

The model is as the following:

Correctness ~ CMPitch + Locative (1 | Participant.no) + (1 | item)

The models, the one with the factor "CMPitch" and the one with the factors "CMPitch" and "Locative" were compared. In this comparison, mean pitch was controlling factor while locative pronoun was tested. Locative did not contribute to the model included only mean pitch ($\chi^2(2) = 0.9305$, $p = 0.628$). Hence, locative did not take part in the model as one of the predictors.

The model continued to be constructed with the fixed structure. The effect of modality on correctness (dependent variable) was analysed in the direction of the following question: Whether modality was the predictor of correctness?

A model was created that included the fixed effect "Modality" in addition to the fixed effect "CMPitch" to predict correctness, controlling for by-participant and by-item variability.

The model is as the following:

Correctness ~ CMPitch + Modality (1 | Participant.no) + (1 | item)

The models, the one with the factor "CMPitch" and the one with the factors "CMPitch" and "Modality" were compared. In this comparison, mean pitch was controlling factor while modality was tested. Modality contributed to the model included only mean pitch ($\chi^2(1) = 25.978$, $p = 3.454\text{e-}07$). Hence, modality took part in the model as one of the predictors.

As a last step, to finalize constructing the model with the fixed structure, interaction between the fixed factors was controlled. The models, the one with the factors "CMPitch" and "Modality" and the one with the interaction between the factors "CMPitch" and "Modality" were compared using anova ( ) function. The result did not yield a statistically significant interaction between mean pitch and modality ($\chi^2(1) = 6e\text{-}04$, $p = 0.98$). Hence, the interaction did not take part in the model.

Selected fixed structure model is as the following:

Correctness ~ CMPitch + Modality + (1 | Participant.no) + (1 | item)

Since the construction of the model with the fixed structure was finalized, it continued with the random structure. The model that was shown above is the random intercept model. However, the effects of mean pitch and modality are not the same for all participants. Accordingly, the variables on the fixed structure were inserted into participant.no slope yet for the items these covariates (independent variables) were not changing. Therefore, item was taken as intercept only. Random structure included the fixed effects "CMPitch" and "Modality" as random slope inserted into participant.no, respectively (i.e., (0 + CMPitch | Participant.no) and (0 + Modality | Participant.no)). The results did not yield statistically significant differences between random intercept and random slope models ($\chi^2(1) = 0$, $p = 1$ and model failed to converge, respectively).

At the end of the fixed and random structuring the selected model is as the following:

Formula: Correctness ~ CMPitch + Modality + (1 | Participant.no) + (1 | item)

Block II

Speaker Confidence (Perceived Modality)

The intercept only model was created by adding random intercepts for participants and items to predict speaker confidence, controlling for by-participant and by-item variability.

Intercept only model is as the following:

Speaker_Confidence ~ 1 + (1 | Participant.no) + (1 | item)

The model started to be constructed with the fixed structure. The effects of the acoustic properties on speaker confidence (dependent variable) were analysed in the direction of the following question: Whether acoustic properties were predictors of speaker confidence?

A model was created that included the fixed effect "CMPitch" (centered mean pitch in Hertz) to predict speaker confidence, controlling for by-participant and by-item variability.

The model is as the following:

Speaker_Confidence ~ CMPitch + (1 | Participant.no) + (1 | item)

The two models, the null and the one with the factor "CMPitch", were compared by performing the likelihood ratio test. The anova ( ) function did not yield a statistically significant difference between these two models. Mean pitch was found ineffective on speaker confidence ($\chi^2(1) = 0.8609$, $p = 0.3535$) and not included in the model.

Another model was created that included the fixed effect "C.Dur" (centered duration in milliseconds) to predict speaker confidence, controlling for by-participant and by-item variability.

The model is as the following:

Speaker_Confidence ~ C.Dur + (1 | Participant.no) + (1 | item)

The two models, the null and the one with the factor "C.Dur", were compared by performing the likelihood ratio test. The anova ( ) function did not yield a statistically significant difference between these two models. Duration was found ineffective on speaker confidence ($\chi^2(1) = 0.3221$, $p = 0.5704$) and not included in the model.

Another model was created that included the fixed effect "CPRange" (centered pitch range in Hertz) to predict speaker confidence, controlling for by-participant and by-item variability.

The model is as the following:

Speaker_Confidence ~ CPRange + (1 | Participant.no) + (1 | item)

The two models, the null and the one with the factor "CPRange", were compared by performing the likelihood ratio test. The anova ( ) function did not yield a statistically significant difference between these two models. Pitch range was found ineffective on correctness ($\chi^2(1) = 0.0214$, $p = 0.8838$) and not included in the model.

The model is as the following:

Speaker_Confidence ~ 1 + (1 | Participant.no) + (1 | item)

The model continued to be constructed with the fixed structure. The effect of modality on speaker confidence (dependent variable) was analysed in the direction of the following question: Whether modality was the predictor of speaker confidence?

A model was created that included the fixed effect "Modality" in addition to the intercept only model to predict speaker confidence, controlling for by-participant and by-item variability.

The model is as the following:

Speaker_Confidence ~ Modality (1|Participant.no) + (1|item)

The models, the intercept only model and the one with the factor "Modality" were compared. In this comparison, modality was tested. Modality contributed to the intercept only model ($\chi^2(1) = 9.5611$, $p = 0.001987$). Hence, modality took part in the model as one of the predictors.

Moreover, the model was compared to the models with the combinations of acoustic properties. The one with the factor "Modality" was compared to the models; the one with the factors "Modality" and "CMPitch", the one with the factors "Modality" and "C.Dur", and the one with the factors "Modality" and "CPRange". According to the results of the comparison of the models, among acoustic properties none of them contributed to the model including only modality (($\chi^2(1) = 1.1542$, $p = 0.2827$), ($\chi^2(1) = 0.2356$, $p = 0.6274$), and ($\chi^2(1) = 0.0899$, $p = 0.7642$), respectively).

Selected fixed structure model is as the following:

Speaker_Confidence ~ Modality + (1 | Participant.no) + (1 | item)

Since the construction of the model with the fixed structure was finalized, it continued with the random structure. The model that was shown above is the random intercept model. However, the effects of modality are not the same for all participants. Accordingly, the variable on the fixed structure was inserted into participant.no slope yet for the items this covariate (independent variables) did not change. Therefore, item was taken as intercept only.

Random structure included the fixed effect "Modality" as random slope inserted into participant.no, (i.e., (0 + Modality | Participant.no)). The model was unidentifiable. Hence, modality did not take part in the model as random slope.

At the end of the fixed and random structuring the selected model is as the following:

Formula: Speaker_Confidence ~ Modality + (1 | Participant.no) + (1 | item)

Correctness (Accuracy)

The intercept only model was created by adding random intercepts for participants and items to predict correctness, controlling for by-participant and by-item variability.

Intercept only model is as the following:

Correctness ~ 1 + (1 | Participant.no) + (1 | item)

The model started to be constructed with the fixed structure. The effects of the acoustic properties on correctness (dependent variable) were analysed in the direction of the following question: Whether acoustic properties were predictors of correctness?

A model was created that included the fixed effect "CMPitch" (centered mean pitch in Hertz) to predict correctness, controlling for by-participant and by-item variability.

The model is as the following:

Correctness ~ CMPitch + (1 | Participant.no) + (1 | item)

The two models, the null and the one with the factor "CMPitch", were compared by performing the likelihood ratio test. The anova ( ) function yielded a statistically significant difference between these two models. Mean pitch was found effective on correctness ($\chi^2(1) = 4.1545$, $p = 0.04153$) and kept in the model.

Another model was created that included the fixed effect "C.Dur" (centered duration in milliseconds) to predict correctness, controlling for by-participant and by-item variability.

The model is as the following:

Correctness ~ C.Dur + (1 | Participant.no) + (1 | item)

The two models, the null and the one with the factor "C.Dur", were compared by performing the likelihood ratio test. The anova ( ) function did not yield a statistically significant difference between these two models. Duration was found ineffective on correctness ($\chi^2(1) = 0.9988$, $p = 0.3176$) and not included in the model.

Another model was created that included the fixed effect "CPRange" (centered pitch range in Hertz) to predict correctness, controlling for by-participant and by-item variability.

The model is as the following:

Correctness ~ CPRange + (1 | Participant.no) + (1 | item)

The two models, the null and the one with the factor "CPRange", were compared performing the likelihood ratio test. The anova ( ) function did not yield a statistically significant difference between these two models. Pitch range was found ineffective on correctness ($\chi^2(1) = 0.663$, $p = 0.4155$) and not included in the model.

Moreover, the models, the one with the factor "CMPitch" and the one with the factors "CMPitch" and "C.Dur" were compared. In this comparison, mean pitch was controlling factor while duration was tested. Duration did not contribute to the model included only mean pitch ($\chi^2(1) = 0.0865$, $p = 0.7686$). Also, the models, the one with the factor "CMPitch" and the one with the factors "CMPitch" and "CPRange" were compared. In this comparison, mean pitch was controlling factor while pitch range was tested. Pitch range did not contribute to the model included only mean pitch ($\chi^2(1) = 0.147$, $p = 0.7014$).

According to the results of the comparison of the models, among acoustic properties mean pitch was found as predictor of correctness and kept in the model.

The model is as the following:

Correctness ~ CMPitch + (1 | Participant.no) + (1 | item)

The model continued to be constructed with the fixed structure. The effect of modality on correctness (dependent variable) was analysed in the direction of the following question: Whether modality was the predictor of correctness?

A model was created that included the fixed effect "Modality" in addition to the fixed effect "CMPitch" to predict correctness, controlling for by-participant and by-item variability.

The model is as the following:

Correctness ~ CMPitch + Modality + (1 | Participant.no) + (1 | item)

The models, the one with the factor "CMPitch" and the one with the factors "CMPitch" and "Modality" were compared. In this comparison, mean pitch was controlling factor while modality was tested. Modality did not contribute to the model included only mean pitch ($\chi^2(1) = 3.4353$, $p = 0.06382$). Hence, modality did not take part in the model as one of the predictors.

Selected fixed structure model is as the following:

Correctness ~ CMPitch + (1 | Participant.no) + (1 | item)

Since the construction of the model with the fixed structure was finalized, it continued with the random structure. The model that was shown above is the random intercept

model. However, the effects of mean pitch and modality are not the same for all participants. Accordingly, the variables on the fixed structure were inserted into participant.no slope yet for the items these covariates (independent variables) were not changing. Therefore, item was taken as intercept only. Random structure included the fixed effect "CMPitch" as random slope inserted into participant.no (i.e., (0 + CMPitch | Participant.no)). The result did not yield statistically significant difference between random intercept and random slope models ($\chi^2(1) = 2.4756$, $p = 0.1156$)

At the end of the fixed and random structuring the selected model is as the following:

Formula: Correctness ~ CMPitch + (1 | Participant.no) + (1 | item)

Block III

Speaker Confidence (Perceived Modality)

The intercept only model was created by adding random intercepts for participants and items to predict speaker confidence, controlling for by-participant and by-item variability.

Intercept only model is as the following:

Speaker_Confidence ~ 1 + (1 | Participant.no) + (1 | item)

The model started to be constructed with the fixed structure. The effects of the acoustic properties on speaker confidence (dependent variable) were analysed in the direction of the following question: Whether acoustic properties were predictors of speaker confidence?

A model was created that included the fixed effect "CMPitch" (centered mean pitch in Hertz) to predict speaker confidence, controlling for by-participant and by-item variability.

The model is as the following:

Speaker_Confidence ~ CMPitch + (1 | Participant.no) + (1 | item)

The two models, the null and the one with the factor "CMPitch", were compared performing the likelihood ratio test. The anova ( ) function did yield a statistically significant difference between these two models. Mean pitch was found effective on speaker confidence ($\chi^2(1) = 7.2737$, $p = 0.006997$) and included in the model.

Another model was created that included the fixed effect "C.Dur" (centered duration in milliseconds) to predict speaker confidence, controlling for by-participant and by-item variability.

The model is as the following:

Speaker_Confidence ~ C.Dur + (1 | Participant.no) + (1 | item)

The two models, the null and the one with the factor "C.Dur", were compared by performing the likelihood ratio test. The anova ( ) function yielded a statistically significant difference between these two models. Duration was found effective on speaker confidence ($\chi^2(1) = 8.159$, $p = 0.004285$) and included in the model.

Another model was created that included the fixed effect "CPRange" (centered pitch range in Hertz) to predict speaker confidence, controlling for by-participant and by-item variability.

The model is as the following:

Speaker_Confidence ~ CPRange + (1 | Participant.no) + (1 | item)

The two models, the null and the one with the factor "CPRange", were compared by performing the likelihood ratio test. The anova ( ) function did not yield a statistically significant difference between these two models. Pitch range was found ineffective on correctness ($\chi^2(1) = 1.9786$, $p = 0.1595$) and not included in the model.

Moreover, the models with the combinations of acoustic properties were also compared. The models, the one with the factor "CMPitch" and the one with the factors "CMPitch" and "C.Dur" were compared. In this comparison, mean pitch was controlling factor while duration was tested. Duration contributed to the model included only mean pitch ($\chi^2(1) = 6.3764$, $p = 0.01157$). Hence, "duration" took part in the model as one of the predictors. In line with this result, while duration was controlling factor and mean pitch was tested, it contributed to the model included only duration ($\chi^2(1) = 5.4911$, $p = 0.01911$).

The model continued to be constructed with the fixed structure. The effect of modality on speaker confidence (dependent variable) was analysed in the direction of the following question: Whether modality was the predictor of speaker confidence?

A model was created that included the fixed effect "Modality" in addition to the fixed effects "CMPitch" and "C.Dur" to predict speaker confidence, controlling for by-participant and by-item variability.

The model is as the following:

Speaker_Confidence ~ CMPitch + C.Dur + Modality + (1 | Participant.no) + (1 | item)

The models, the one with the fixed effects "CMPitch" and "CDur", and the one with the factor "Modality" were compared. In this comparison, while mean pitch and duration

were controlling factors, modality was tested. Modality contributed to the model ($\chi 2(1) = 3.0778$, $p = 0.07937$). Hence, modality did not take part in th model as a predictor.

Selected fixed structure model is as the following:

Speaker_Confidence ~ CMPitch + C.Dur + (1 | Participant.no) + (1 | item)

Moreover, for interactions between "CMPitch" * "C.Dur" model failed to converge.

Since the construction of the model with the fixed structure was finalized, it continued with the random structure. The model that was shown above is the random intercept model. However, the effect of mean pitch and duration are not the same for all participants. Accordingly, the variable on the fixed structure was inserted into participant.no slope yet for the items this covariate (independent variable) was not changing. Therefore, item was taken as intercept only. Random structure included the fixed effect "C.Dur" and "CMPitch" as random slope inserted into participant.no., respectively (i.e., (0 + CMPitch | Participant.no), (0 + C.Dur | Participant.no)). The results did not yield statistically significant differences between random intercept and random slope models (($\chi^2(1) = 0$, $p = 1$) and ($\chi^2(1) = 0.0617$, $p = 0.8038$), respectively).

At the end of the fixed and random structuring the selected model is as the following:

Formula = Speaker_Confidence ~ CMPitch + C.Dur + (1 | Participant.no) + (1 | item)

Correctness (Accuracy)

The intercept only model was created by adding random intercepts for participants and items to predict correctness, controlling for by-participant and by-item variability.

Intercept only model is as the following:

Correctness ~ 1 + (1 | Participant.no) + (1 | item)

The model started to be constructed with the fixed structure. The effects of the acoustic properties on correctness (dependent variable) were analysed in the direction of the following question: Whether acoustic properties were predictors of correctness?

A model was created that included the fixed effect "CMPitch" (centered mean pitch in Hertz) to predict correctness, controlling for by-participant and by-item variability.

The model is as the following:

Correctness ~ CMPitch + (1 | Participant.no) + (1 | item)
The two models, the null and the one with the factor "CMPitch", were compared by performing the likelihood ratio test. The anova ( ) function did not yield a statistically

significant difference between these two models. Mean pitch was found ineffective on correctness ($\chi^2(1) = 0.0011$, $p = 0.9739$) and not included in the model.

Another model was created that included the fixed effect "C.Dur" (centered duration in milliseconds) to predict correctness, controlling for by-participant and by-item variability.

The model is as the following:

Correctness ~ C.Dur + (1 | Participant.no) + (1 | item)

The two models, the null and the one with the factor "C.Dur", were compared by performing the likelihood ratio test. The anova ( ) function did not yield a statistically significant difference between these two models. Duration was found ineffective on correctness ($\chi^2(1) = 2.4796$, $p = 0.1153$) and not included in the model.

Another model was created that included the fixed effect "CPRange" (centered pitch range in Hertz) to predict correctness, controlling for by-participant and by-item variability.

The model is as the following:

Correctness ~ CPRange + (1 | Participant.no) + (1 | item)

The two models, the null and the one with the factor "CPRange", were compared by performing the likelihood ratio test. The anova ( ) function did not yield a statistically significant difference between these two models. Pitch range was found ineffective on correctness ($\chi^2(1) = 0.1103$, $p = 0.7398$) and not included in the model.

The model continued to be constructed with the fixed structure. The effect of modality on correctness (dependent variable) was analysed in the direction of the following question: Whether modality was the predictor of correctness?

A model was created that included the fixed effect "Modality" to predict correctness, controlling for by-participant and by-item variability.

The model is as the following:

Correctness ~ Modality + (1 | Participant.no) + (1 | item)

The models, the intercept only model and the one with the factor "Modality" were compared. In this comparison, modality was tested. Modality did not contribute to the intercept only model ($\chi^2(1) = 0.0057$, $p = 0.9401$). Hence, modality did not take part in the model as one of the predictors.

Neither acoustic features nor modality predicted correctness.

Selected fixed structure model is as the following:

Formula: Correctness ~ 1 + (1 | Participant.no) + (1 | item)

Block IV.I

Speaker Confidence Score (Perceived Modality Score)

The intercept only model was created by adding random intercepts for participants and items to predict speaker confidence score for certainty, controlling for by-participant and by-item variability.

Intercept only model is as the following:

Speaker_Confidence_Score ~ 1 + (1 | Participant.no) + (1 | item)

The model started to be constructed with the fixed structure. The effects of the acoustic properties on speaker confidence score for certainty (dependent variable) were analysed in the direction of the following question: Whether acoustic properties were the predictors of speaker confidence score for certainty?

A model was created that included the fixed effect "CMPitch" (centered mean pitch in Hertz) to predict speaker confidence score for certainty, controlling for by-participant and by-item variability.

Speaker_Confidence_Score ~ CMPitch + (1 | Participant.no) + (1 | item)

The two models, the null and the one with the factor "CMPitch", were compared by performing the likelihood ratio test. The anova ( ) function did not yield a statistically significant difference between these two models. Mean pitch was found ineffective on speaker confidence score for certainty ($\chi^2(1) = 0.6579$, $p = 0.4173$) and not included in the model.

Another model was created that included the fixed effect "C.Dur" (centered duration in milliseconds) to predict speaker confidence score for certainty, controlling for by-participant and by-item variability.

The model is as the following:

Speaker_Confidence_Score ~ C.Dur + (1 | Participant.no) + (1 | item)

The two models, the null and the one with the factor "C.Dur", were compared by performing the likelihood ratio test. The anova ( ) function yielded a statistically

significant difference between these two models. Duration was found effective on speaker confidence score for certainty ($\chi^2(1) = 5.4779$, $p = 0.01926$) and kept in the model.

Another model was created that included the fixed effect "CPRange" (centered pitch range in Hertz) to predict speaker confidence score for certainty, controlling for by-participant and by-item variability.

The model is as the following:

Speaker_Confidence_Score ~ CPRange + (1 | Participant.no) + (1 | item)

The two models, the null and the one with the factor "CPRange", were compared by performing the likelihood ratio test. The anova ( ) function did not yield a statistically significant difference between these two models. Pitch range was found ineffective alone on speaker confidence score for certainty ($\chi^2(1) = 3.766$, $p = 0.05231$) and not included in the model.

Moreover, the models, the one with the factor "C.Dur" and the one with the factors "C.Dur" and "CPRange" were compared. In this comparison, duration was controlling factor while pitch range was tested. Pitch range contributed to the model included only duration ($\chi^2(1) = 7.3371$, $p = 0.006755$). Furthermore, the models, the one with the factors "C.Dur" and "CPRange" and the one with the factors "C.Dur", "CPRange", and "CMPitch" were compared. In this comparison, duration and pitch range were controlling factors while mean pitch was tested. Mean pitch did not contribute to the model included duration and pitch range ($\chi^2(1) = 0.0247$, $p = 0.875$). According to the results of the comparison of the models, among acoustic properties duration and pitch range were found as predictors of speaker confidence score for certainty and kept in the model.

The model is as the following:

Speaker_Confidence_Score ~ C.Dur + CPRange + (1 | Participant.no) + (1 | item)

The interaction between the fixed factors was controlled. The models, the one with the factors (Speaker_Confidence ~ C.Dur + CPRange + (1 | Participant.no) + (1 | item)) and the one with the interaction between the factors (Speaker_Confidence ~ C.Dur * CPRange + (1 | Participant.no) + (1 | item)) were compared using anova ( ) function. The model failed to converge. Hence, the interaction did not take part in the model.

The model continued to be constructed with the fixed structure. The effect of the locative pronouns on speaker confidence score (dependent variable) was analysed in the direction of the following question: Whether locative was the predictor of speaker confidence score for certainty?

A model was created that included the fixed effect of "Locative" in addition to the fixed effects of "C.Dur" and "CPRange" to predict speaker confidence score for certainty, controlling for by-participant and by-item variability.

The model is as the following:

Speaker_Confidence_Score ~ C.Dur + CPRange+ Locative + (1 | Participant.no) + (1 | item)

The models, the one with the factors "C.Dur" and "CPRange", and the one with the factors "C.Dur", "CPRange", and "Locative" were compared. In this comparison, duration and pitch range were controlling factors while locative pronouns were tested. Locative pronouns did not contribute to the model ($\chi^2(2) = 1.2946$, $p = 0.5235$). Hence, locative did not take part in the model as one of the predictors.

Selected fixed structure model is as the following:

Speaker_Confidence_Score ~ C.Dur + CPRange + (1 | Participant.no) + (1 | item)

Since the construction of the model with the fixed structure was finalized, it continued with the random structure. The model that was shown above is the random intercept model. However, the effects of duration and pitch range are not the same for all participants. Accordingly, the variables on the fixed structure were inserted into participant.no slope yet for the items these covariates (independent variables) were not changing. Therefore, item was taken as intercept only. Random structure included the fixed effect "C.Dur" and "CPRange" as random slopes inserted into participant.no, respectively (i.e., (0 + C.Dur | Participant.no) and (0 + CPRange | Participant.no)). The models failed to converge. These fixed effects could not be used in the model as random slopes.

At the end of the fixed and random structuring the selected model is as the following:

Formula: Speaker_Confidence_Score ~ C.Dur + CPRange + (1 | Participant.no) + (1 | item)

Block IV.II

Speaker Confidence Score (Perceived Modality Score)

The intercept only model was created by adding random intercepts for participants and items to predict speaker confidence score for uncertainty, controlling for by-participant and by-item variability.

Intercept only model is as the following:

Speaker_Confidence_Score ~ 1 + (1 | Participant.no) + (1 | item)

The model started to be constructed with the fixed structure. The effects of the acoustic properties on speaker confidence score for uncertainty (dependent variable) were analysed in the direction of the following question: Whether acoustic properties were the predictors of speaker confidence score for uncertainty?

A model was created that included the fixed effect "CMPitch" (centered mean pitch in Hertz) to predict speaker confidence score for uncertainty, controlling for by-participant and by-item variability.

Speaker_Confidence_Score ~ CMPitch + (1 | Participant.no) + (1 | item)

The two models, the null and the one with the factor "CMPitch", were compared by performing the likelihood ratio test. The anova ( ) function did not yield a statistically significant difference between these two models. Mean pitch was found ineffective on speaker confidence score for uncertainty ($\chi^2(1) = 0.5119$, $p = 0.4743$) and not included in the model.

Another model was created that included the fixed effect "C.Dur" (centered duration in milliseconds) to predict speaker confidence score for uncertainty, controlling for by-participant and by-item variability.

The model is as the following:

Speaker_Confidence_Score ~ C.Dur + (1 | Participant.no) + (1 | item)

The two models, the null and the one with the factor "C.Dur", were compared by performing the likelihood ratio test. The anova ( ) function yielded a statistically significant difference between these two models. Duration was found effective on speaker confidence score for certainty ($\chi2(1) = 19.406$, p = 1.057e-05) and kept in the model.

Another model was created that included the fixed effect "CPRange" (centered pitch range in Hertz) to predict speaker confidence score for uncertainty, controlling for by-participant and by-item variability.

The model is as the following:

Speaker_Confidence_Score ~ CPRange + (1 | Participant.no) + (1 | item)

The two models, the null and the one with the factor "CPRange", were compared by performing the likelihood ratio test. The anova ( ) function did not yield a statistically significant difference between these two models. Pitch range was found ineffective

alone on speaker confidence score for uncertainty ($\chi^2(1) = 1.2049$, $p = 0.2724$) and not included in the model.

Moreover, the models, the one with the factor "C.Dur" and the one with the factors "C.Dur" and "CMPitch" were compared. In this comparison, duration was controlling factor while mean pitch was tested. Mean pitch did not contribute to the model included only duration ($\chi^2(1) = 1.7095$, $p = 0.1911$). Also, the models, the one with the factor "C.Dur" and the one with the factors "C.Dur" and "CPRange" were compared. In this comparison, duration was controlling factor while pitch range was tested. Pitch range did not contribute to the model included only duration ($\chi^2(1) = 1.7837$, $p = 0.1817$). Furthermore, the models, the one with the factor "C.Dur" and the one with the factors "C.Dur", "CMPitch", and "CPRange" were compared. In this comparison, duration was controlling factor while mean pitch and pitch range were tested. They did not contribute to the model included only duration ($\chi^2(2) = 4.561$, $p = 0.1022$). According to the results of the comparison of the models, among acoustic properties duration was found as a predictor of speaker confidence for uncertainty and kept in the model.

The model is as the following:

Speaker_Confidence_Score ~ C.Dur + (1 | Participant.no) + (1 | item)

The model continued to be constructed with the fixed structure. The effect of the locative pronouns on speaker confidence score (dependent variable) was analysed in the direction of the following question: Whether locative was the predictor of speaker confidence score for uncertainty?

A model was created that included the fixed effect "Locative" in addition to the fixed effect "C.Dur" to predict speaker confidence score for certainty, controlling for by-participant and by-item variability.

The model is as the following:

Speaker_Confidence_Score ~ C.Dur + Locative (1 | Participant.no) + (1 | item)

The models, the one with the factor "C.Dur" and the one with the factors "C.Dur" and "Locative" were compared. In this comparison, duration was controlling factor while locative pronouns were tested. Locative pronouns did not contribute to the model included only duration ($\chi^2(2) = 0.1168$, $p = 0.9433$). Hence, locative did not take part in the model as one of the predictors.

Selected fixed structure model is as the following:

Speaker_Confidence_Score ~ C.Dur + (1 | Participant.no) + (1 | item)

Since the construction of the model with the fixed structure was finalized, it continued with the random structure. The model that was shown above is the random intercept model. However, the effect of duration is not the same for all participants. Accordingly, the variable on the fixed structure were inserted into participant.no slope yet for the items this covariate (independent variables) was not changing. Therefore, item was taken as intercept only. Random structure included the fixed effect "C.Dur" as random slope inserted into participant.no (i.e., (0 + C.Dur | Participant.no)). The models failed to converge. This fixed effect could not be used in the model as random slope.

At the end of the fixed and random structuring the selected model is as the following:

Formula: Speaker_Confidence_Score ~ C.Dur + (1 | Participant.no) + (1 | item)