Prosodic encoding of information structure by Greek speakers with Autism

Dafni Vaia Bagioka

Master's Thesis

Leiden University

Leiden University Centre for Linguistics

Faculty of Humanities

December 2020

Supervisor: Dr. Yiya Chen

Second Reader: Dr. Marina Terkourafi

Contents

Acknowledgements

I would like to express my greatest appreciation to Yiya Chen. Completing this thesis under her supervision has been an enriching experience. I have learnt so many new things through our interaction and, most importantly, I had the opportunity to see how theory is put to use in an experiment. I could not omit to thank the lecturers of the courses I attended in the master's program. Marina Terkourafi introduced me to the world of Pragmatics. Andreea Geambaşu taught me all the necessary methodological and ethical issues that should be considered before the beginning of research. Leticia Pablos Robles and Jurriaan Witteman offered me advanced knowledge of statistics and R coding. Rint Sybesma made me aware of the latest developments in syntactic theory. It is important to mention that, as a speech therapist, studying Linguists seemed an interesting challenge for me and the knowledge that I gained is indisputably beneficial for my future academic and professional career. Studying at Leiden University has been a wonderful experience and it was a great environment for expanding my international network.

I am deeply indebted to all the people in Greece, without whom I would not have been able to run this experiment. My special thanks go to Arhonto Terzi, who supported me with wise and valuable advice throughout my master's studies, and Voula Georgopoulos for her supply of the recording equipment. I would like to extend my sincere thanks to Michaela Nerantzini and Ioannis Papakyritsis for their important support at the setup of the experiment. I also wish to thank the psychiatrists Konstantinos Bolias and Panagiotis Vutirakos for their help with participant recruitment. Most importantly, I would like to express my gratitude to all the participants, their families and the therapists that kindly agreed to take part in the study. I cannot begin to express my thanks to my linguist friends, Antonis Panagiotopoulos and Carolina Lins-Machado, who gave me feedback through the entire process of the thesis writing and my friends Ioanna, Nicole and Giorgos who voluntarily participated in the photo shoot for the creation of the visual stimuli used in the experiment. Finally, I am extremely grateful to my parents, Grigoris and Kyriaki, who provided me with strength when I needed it the most. This thesis is dedicated to you.

Abstract

One aspect that heavily impacts the social communication of people with Autism Spectrum Disorder (ASD) is abnormal prosody. Previous studies have shown heterogeneity on acoustic properties of the prosodic encoding of information structure phenomena even in High-Functioning individuals with Autism (HFA). The main goal of this study is to investigate whether and if so, how Greek HFA speakers with different language abilities mark information structure, in comparison to neurotypical speakers. Thirty native speakers of Greek (16-27 years old) took part in a question-answer task, in which the F0 and duration of the subject and the object of their productions were measured in different focus conditions at word and syllable level. Acoustic analyses revealed that HFA participants exhibited longer duration at word level and that in most of the group differences, HFA with moderate language skills were found to differ more often from the neurotypical group than the HFA with high language skills. Thus, the level of language skills of HFAs does correlate with their ability to encode information structure with prosody. Results of this study suggest that it is important for speech therapists to first understand the details of prosodic use in different HFA subgroups before applying any remediation strategy.

List of Tables

List of Figures

Introduction

Autism is a neurodevelopmental disorder characterized by failure to initiate or respond to social interactions and communication. It is associated with repetitive, restricted, stereotyped behavior, interests, and/or activities [DSM-V (American Psychiatric Association, 2013); ICD-10 (World Health Organization, 2019)]. The spectrum ranges from individuals with "low-functioning'' autism to "high-functioning'' autonomous individuals. The former is associated with difficulties which make them dependent on someone else (e.g., caregiver). The latter seems to have a normal everyday life but in reality, they often face problems in social communication creating difficulties in workplace and personal relationships.

One aspect that heavily impacts the social communication of people with Autism Spectrum Disorder (ASD) is impaired prosody (i.e. atypical acoustic patterns). Prosody has been found to be atypical in individuals with ASD cross-linguistically. This includes inappropriate accent placement (Shriberg et al., 2001) and a difficulty of the appropriate usage of pitch or/and duration in marking information structure (DePape et al., 2012). This thesis investigates the prosodic encoding of information structure of high functioning individuals with ASD. The main goal of this study is to answer whether Greek speakers with ASD have difficulties in marking information structure through their voices. In addition, the findings may offer an explanation as to whether individuals with ASD face problems in social communication because of this difficulty in prosodic encoding of information structure.

Information structure, here, refers to the way information is packaged within a sentence through prosody and the presence or the absence of focus [see e.g., Krifka (2008) for more discussion on notions of information structure]. Depending on the focus condition, the same sentence can convey different kinds of information to the listener. There are at least three ways in which the information structure of a sentence element can be conveyed: using a) word order, b) particular lexical items and syntactic constructions, and c) prosody. This thesis will only deal with the prosody used in focus phenomenon.

Focus refers to the marking of new information in an utterance by placing prosodic prominence on it. Given (or old) information is already known by the listener and does not receive focus marking. Consider the following questions (1a, 1b, 1c), which always have as an answer the same sentence (Mary cut the yarn), but with prosodic prominence on a different element in each case:

(1) a. What happened?

[Mary cut the yarn.]

b. Who cut the yarn?

[MARY] cut the yarn.

c. What did Mary cut?

Mary cut the [YARN].

In the above examples, two dimensions of focus can be observed, the focus breadth and the focus location. Focus breadth is a basic dimension along which focused elements can vary (Bosch & Sandt, 1999; Goldsmith et al., 1995; Skopeteas, 2016). Focus breadth refers to the size of the set of focused elements and it consists of two categories: narrow and wide (Gussenhoven, 1999; Selkirk, 1995). For instance, wide focus refers to the case where the entire event is focused, such as the answer to the first question ("Mary cut the yarn"). In contrast, narrow focus refers to cases where only a specific aspect of an event is focused, such as the sentences that are the replies to questions (1b): "Mary" and (1c): "yarn". Based on the location of narrowly focused elements have in the sentence, a new dimension of focus is created i.e., the focus location. In the above examples, the focus location has two subcategories: Subject-"Mary" and Object-"the yarn", corresponding to questions (1b) and (1c).

For the current study, the questions (1a) - (1c) have been used to elicit spontaneous but structured speech. This question-answer task retains the benefits of natural conversation while still providing controlled, quantifiable, and comparable acoustic productions. The acoustic realization of focus in Greek utterances will be investigated through pitch and duration measures. The thesis is structured as follows: Chapter 1 discusses the prosodic characteristics via which information structure is encoded in different languages with particular attention to Greek and to how people with ASD use these features. Chapter 2 and Chapter 3 offer a detailed presentation of the production study and the results, respectively. In Chapter 4, I discuss the findings.

Chapter 1
Literature Review

1.1 Information Structure Cross linguistically

Within cross-linguistic studies, information structure has been studied through different methods and frameworks. There are two general perspectives on the relationship between the acoustics of the speech signal and the meaning that is associated with various aspects of focus. First, some researchers investigate sets of acoustic features, which are directly associated with particular meanings following the direct relationship approach (Breen et al., 2010). In contrast, some others study the relationship between acoustics and meaning, which is mediated by phonological categories, according to the intonational phonology framework (Arvaniti & Baltazani, 2005; Dilley, 2005; Ladd, 2008). In other words, the latter is an indirect-relationship approach, in which it is argued that phonetic prosodic cues serve to signal the prosodic categories which are associated with particular meanings.

The experiment in this study is closer to the first approach (direct relationship approach) as focus is studied through the acoustic features of the F0 and duration, but it was not designed to argue in favor of it. To be more specific, the aim of this study is to investigate the direct acoustic correlates of focus, which can shed some light on the indirect mapping approach of it.

Having the same approach, Breen et al. (2010) demonstrate, through three experiments in English, that duration, mean F0, maximum F0 and maximum intensity are the four (4) – out of the 24 candidate acoustic features that they examined— most important features which encode differences among focus conditions. Moreover, Rao et al. (2017) show that duration is probably the most significant acoustic feature in the distinction of focus conditions in Marathi. Characteristically, they stated that “Only when duration is ambiguous, does the on-focus F0 cue appear to play a role” (p.1). These results indicate that one or more acoustic features play the most vital role in the acoustic realization of prominence in the focused element among languages; F0 and duration are typically the measurements of interest in most of the studies, including studies on Greek, as it will be shown in the following section.

1.2 Information Structure in Greek

It is important to take into account that Greece is not a country with strong phonetic tradition (Arvaniti, 2007). Nevertheless, from the beginning of the 21st century and during the last two decades the body of research on Greek phonetics has done remarkable work in trying to fill in the gaps in phonetics and especially in research on prosodic encoding of information structure.

In earlier studies (Baltazani & Jun, 1999; Botinis & Bannert, 2003) in Greek, the researchers investigated information structure and focus on different phenomena and this is, probably, a reason for which they present mixed results. For instance, Baltazani and Jun (1999) argued that focus lengthens the stressed[1] syllable and therefore the focused word. On the other hand, Botinis and Bannert (2003) proposed that focus affects only the duration of the stressed syllable. Thus, in earlier studies, there was no robust evidence to suggest if the narrow focus that is related to information structure is manifested in the same or in a different way from focus on the whole word.

Furthermore, Baltazani (2003) reports that H*[2] signals broad focus and in more recent studies, focus has been examined through investigating pitch accents. In this view, Arvaniti et al. (2006) support the opinion that the L+H* nuclear accent signals narrow or contrastive focus and they state that durational differences across utterance types are observed. Also, Arvaniti & Baltazani (2005) propose that there are two possible nuclear accents in broad focus declaratives, which they represented as H* and H*+L. Nowadays, Lohfink et al. (2019), including more parameters from pragmatics, made a new distinction between the above H* and H*+L. They propose both pitch accents indicate that the accented item is new in discourse, but H*+L additionally indicate that the speaker has a familiarity with the focused item.

Although the above-mentioned studies provide evidence for some differences in the acoustic realization of different aspects of focus, no solid conclusions have yet been reached regarding the acoustic differences between: 1) focused vs. given elements and 2) narrowly vs. widely focused elements. Thus, until this point of the literature review in Greek, it seems that

[1] In Greek the position of primary stress falls on one of the last three syllables of the word. Also, in Greek primary stress cannot be predicted from phonological structure as it may be the case in other languages.

[2] In Greek, based on GRToBI (Arvaniti & Baltazani, 2014), there are five pitch accents: H*, L*, L*+H, L+H*, H* and H*+L. From them, H*, L+H* and H*+L signals narrow or broad focus, in different types of sentences. H* is an accent which "is realized as a peak on the accented syllable, but lacks the initial dip associated with the L". "The realizational difference between L+H* and L*+H lies in the alignment of the H tone: the H tone of L+H* is well within the accented vowel, whereas the H tone of the L*+H aligns early in the first post-accentual vowel".

researchers follow more the indirect approach. Based on this, it is generally agreed that different pitch accents typically express different focus conditions – sometimes there is cross-category overlap (Lohfink et al., 2019) — without specifying anything about the F0 and duration measures in these different conditions. Also, almost all the above studies involve frameworks from phonology and pragmatics, such as Tones and Break Indices (ToBI) and Functional Principal Component Analysis (FPCA). In the current study, I report acoustic features in order to avoid confusion about what the ToBI and FPCA labels might mean and in order not to presuppose the existence of prosodic categories associated with particular meaning categories of information structure.

1.3 Prosody and Information Structure in Autism

There has been no conclusive evidence so far on the acoustic markers for ASD individuals in the literature but, as presented in this section, atypical acoustic patterns of speech have been observed in individuals with ASD.

McCann and Peppé (2003) reviewed sixteen early studies about prosody of individuals with ASD and revealed many significant findings across these studies. However, the three basic limitations identified across all of them were: a) insufficient clinical sample sizes with limited control data, b) no detailed methodology and c) lack of standardized measures. In addition, only two (Baltaxe et al., 1984; Fosnot & Jun, 1999) out of the sixteen studies made acoustic measurements of prosody, while others used clinical valid but subjective measures, from which no clear conclusion can be drawn.

The first findings about prosody and ASD were contradictive. Specifically, Baltaxe et al. (1984) argued that there was no difference between the intonation patters of individuals with ASD and neurotypical controls, while Fosnot and Jun (1999) reported atypical intonation patterns in people with ASD. In the former study, five children with autism (4–12 years old) were compared to children with language impairment and typically developing children matched on language level by having the same mean length in their utterances. In the latter study, four individuals with autism (7-14 years old) were matched on chronological age with typical controls and children who stuttered. Results demonstrated that children with ASD used a wider pitch range and had greater variation in F0 than the others when reading short sentences and imitating sentences produced by the examiner. It should be noted that in both studies there is no information regarding the language abilities and general functionality level of the participants with ASD.

Moving on to the work of the last two decades, the functionality level of participants with ASD is taken into account. By examining the fundamental frequency variation through retelling an emotional story, Edelson et al. (2007) found that the high-functioning individuals with ASD (HFA) had significantly higher pitch and a wider pitch range than the control group. Moreover, Diehl et al. (2009) identified that HFA children and adolescents have greater F0 variation during narrative production compared to their respective control groups. Nadig and Shaw (2012) specified, except for functionality, the language level of the participants. Especially, they matched on age, gender and language level 15 HFA children with 13 typically developing children and found increased F0 range in HFA.

In the above studies, acoustic analyses generally found increased pitch range in children with ASD. In contrast, by investigating imitative productions of nonsense syllables, Paul et al. (2008) found that individuals with ASD produced less lengthening on stressed syllables, as well as a pattern of increased F0 range for both stressed and unstressed syllables compared to a control group. Thus, this was the first indication that further research is needed in order to understand the prosodic use of duration in ASD. As a result, pitch range and variability did not entirely capture the abnormal nature of prosody in those with ASD as it were characterized in perceptual judgments of prosody (DePape et al., 2012).

Fusaroli et al. (2017) systematically reviewed the literature on distinctive acoustic patterns in ASD and could not find a predictor for severity of clinical features, concluding that there is still unsystematic avenue for establishing ASD markers. They explain that this happens because their adopted methods were too diverse. This means that there were many methodological differences between the studies, in order to determine which acoustic feature(s) create(s) these atypical acoustic patterns. Nevertheless, acoustic analysis together with perceptual judgments of prosody, clearly demonstrates that prosody is impacted in ASD (Patel et al., 2020).

Another significant issue is that most studies of prosody in ASD have examined children rather than adults or even adolescents. As far as I am aware, there are only two studies (DePape et al., 2012; Krüger et al., 2018) which focused exclusively on adolescents and adults with HFA. All of them follow different tasks and analysis methods.

Particularly, Krüger et al. (2018), with a cooperative story-telling task in 16 native German adults (25-55 years old) with HFA, discovered a reduced ability in individuals with ASD to mark information structure via pitch. Furthermore, DePape et al. (2012) divided 12 young adults and adults (17-34 years old) with HFA in two groups by the level of their language abilities. Through a question-answer task, they found that the six males with HFA and high language skills use the same pitch range as neurotypical adults, but they do not mark information structure appropriately. However, although the six males with HFA and moderate language skills adults with ASD use smaller pitch range, they mark it appropriately with longer word duration, in the same way as the neurotypical adults do. The advantage of this study was that the researchers take into account the parameter of language abilities, albeit its basic limitation is the size of the sample in combination with the wide age range of the participants.

Unfortunately, in Greek, there is no study which solely focuses on prosody and information structure. On the one hand, studies about children with ASD emphasize on the syntactic and/or the pragmatic domain of the information structure (Marinis et al., 2013; Terzi et al., 2014, 2016). For instance, Terzi et al.(2016) indicate that children with ASD showed atypical understanding of prosody when they had to use a focus structure along with clitics, without reporting any acoustic results. On the other hand, there is a case study in which pitch range and duration are investigated but not in relation to information structure (Tripolitou & Chaida, 2011). In the later study, a female adult (43 years old) with HFA used pitch and duration appropriately in order to mark polar questions and statements. This female participant was compared with a control group of 5 female normal speakers, without further specification of the age of the control participants. Another issue in this study is that it implies that the speakers were reading the sentences. This methodology creates validity issues because tasks of this nature do not elicit natural productions. Furthermore, there is no reference to her language abilities. Therefore, these findings raise the question of the relationship between language abilities and usage of the acoustic cues of the individual with HFA.

1.4 Research Questions and Hypotheses

The current study aimed to build on prior work by comparing prosodic encoding profiles of information structure among adolescents and adults with HFA and their matched – on age, gender, and language abilities – neurotypical group. Overall, I tested the hypothesis that both neurotypical individuals and individuals with HFA use pitch and duration to mark information structure in Greek productions. However, individuals with HFA and the neurotypical group would have differences in the way they use these two cues, therefore differences in one or more acoustic measures. Additionally, because people with HFA could have high or moderate language skills, I divided the participants with ASD in two groups based on this criterion. I expected that acoustic differences from each group would be related to the language abilities between each group. Sex differences were also explored.

The selection of this thesis topic was motivated by several reasons. Firstly, as far as I am aware there is no published work that employs prosodic encoding of information structure on autism in Greek. Therefore, through this thesis I will be able to shed some light on this understudied area. Secondly, acoustic analysis can be a useful tool in order to conduct perceptual studies and clinical assessment of autism. Thus, acoustic studies that look into the deviant speech characteristics of autism in Greek are needed. Thirdly, there are some secondary questions which emerged as a result of a) the literature review and b) the author's clinical experience as a speech pathologist. One of these issues is whether the syllable and the word level analysis of the focus phenomenon gives the same result; this question emerged from the findings of the first acoustic Greek studies (Baltazani & Jun, 1999; Botinis & Bannert, 2003). Another issue that has not been formally addressed in any study, which concerns speech pathologists who work with people with ASD, is whether the differences on language abilities are related to the gender factor. In other words, it is believed informally that females with autism usually have better language abilities than males. Of course, for answering the last two issues additional studies should be conducted, but the findings of the current study should provide a first indication and set the grounds for further research.

Chapter 2
Methods

To test the hypotheses presented in Chapter 1, a production experiment was conducted. The information structure of productions was elicited by manipulating questions that participants had to answer. This section provides a detailed explanation of the methodology behind the experiment. Firstly, it presents the profiles of the participants and the creation of stimuli, followed by the administration of the experiment and the design of the analysis. All stages of the research complied with the Ethics Code for linguistics research set out by the Leiden University Centre for Linguistics.

2.1 Participants

The participants are thirty (30) monolingual native speakers of Greek (18 males) aged from 16 to 27 years old. 20 of them are individuals with ASD (12 males: age M±sd = 21.75+/-4.14 and 8 females: age M±sd = 19.63+/-3.99), who had all been diagnosed with ICD-10: F84.0 (Childhood autism: Autism disorder). They all were recruited from rehabilitation centers in Athens. In addition, 10 neurotypical controls (6 males: age M±sd = 21.17+/-3.60 and 4 females, M±sd = 19.75+/-2.88) were recruited from the experimenter's social circle. None of the participants had any hearing or visual problems. Participants (or the participants' parent/caregiver – for the ones under the age of 18–) signed an informed consent form (see Appendix A) before taking part and were fully debriefed on the intention of the study upon completion of the experiment. Participation in the study was voluntary.

All participants completed the standardized Boston Naming Test (BNT) (Goodglass et al. 1983) as it is adjusted in Greek (see Appendix B). A questionnaire on family history of ASD was given after the experiment to their therapists or to the participants directly if they were neurotypical controls (see Appendix C). Participants were divided in three groups (10 participants/group) depending on their BNT score. Specifically, Group I (High-A) consisted of 6 males and 4 females with ASD, who had the highest BNT score in each gender. Group II (Mod-A) consisted of the rest 6 males and 4 females with ASD and lowest scores in BNT. 9 out of 10 participants in the Mod-A group experienced early language delay whereas none in the High-A group experienced early language delay. Neurotypical participants were the Group III (NT) and none of them had experienced early language delay. None of the participants in NT group had any family member diagnosed with ASD. Furthermore, all participants were matched in age and gender, therefore creating 10 triplets. In other words, for every participant of High-A group there is one in the Mod-A and one in the NT group with the same gender and similar age (± 0-4 years), as shown in Table 1.

Table 1 | Demographic and background information by group

	High-A		NT		Mod-A		
	Age (years)	**BNT**	**Age (years)**	**BNT**	**Age (years)**	**BNT**	
Individuals scores	17	89	17	86	16	68	**Males**
	17	82	16	82	16	62	
	24	81	24	82	25	61	
	26	72	22	82	27	59	
	25	71	26	80	25	57	
	22	71	22	79	22	56	
	16	82	16	81	16	79	**Females**
	24	81	20	81	27	77	
	17	81	19	80	18	76	
	25	80	24	81	25	65	
Mean ± SD	21.3 +/-4.06	79.1 +/-5.90	20.6 +/-3.57	81.40 +/-1.90	21.7 +/-4.72	66.0 +/-8.60	

2.2 Stimuli

In order to make the experimental productions as comparable as possible, the extraction of acoustic features as easy as possible, and to avoid extra semantic and prosodic effects, the following three criteria were taken into account in constructing the materials.

- ✓ Length and stress: All subjects and objects in the target productions were disyllabic words with first syllable stress.
- ✓ Sonority: all words were comprised mostly of sonorant phonemes based on the sonority scale in Papakyritsis et al. (2019).
- ✓ Semantic effect: All subjects were proper names and all objects were mostly common inanimate objects.

Based on these criteria, 13 sets of items are constructed. A sample item is presented in Table 2. The complete set of materials can be found in Appendix D.

Table 2 | A sample item from production experiment

Target sentence:	Η Μέρι [i 'Meri]$_s$ "Mary	έκοψε ['ekopse]$_v$ cut	το νήμα. [to 'nima]$_O$ the yarn.	(Greek)

Condition	**Focus location**	**Set up question**
0	Wide focus	[ti ˈejine eðo] "What happened here?"
1	Narrow focus on Subject	[pços ˈekopse to 'nima] "Who cut the yarn?"
2	Narrow focus on Object	[ti 'ekopse i ˈmeri] "What did Mary cut?"

2.2.1. Randomization

In total, there are 13 question sets with three conditions each: Wide focus (WF), Narrow Focus on Subject (FS) and Narrow Focus on Object (FO). The first set was the same for all the participants in the familiarization phase together with two fillers. During the main task, the remaining 12 sets (36 questions) were presented in three blocks (12 in each block). Each block included randomly only one out of the three conditions of the 12 questions and 6 filler questions (one filler after every 1, 2 or 3 test questions). The blocks were presented in random order. Also, within the three blocks, questions and fillers were presented in random order. Thus, each participant was examined in all the questions in random order. The complete set of materials with the fillers can be found in Appendix D and E.

2.2.2. Creation of stimuli

The stimuli (visual and auditory) were produced by the researcher in three steps. The first step (Auditory stimuli) involved recording the names of the 3 subjects, 8 objects and 12 actions and the 59 questions for the experiment. In the second step (Visual stimuli) photos that depict these objects and actions were taken. The last step (Combination) involved the combination of the photos with the recordings.

Auditory Stimuli: Productions were recorded in a quiet room with a head-mounted microphone at a rate of 44.1 kHz. The experimenter – who is a phonologically trained female native speaker of standard Greek- recorded the words/names and the questions using Audacity (Version 2.3.3).[3]

Visual Stimuli: Real-life objects and subjects were photographed by the researcher. Photos were edited using Adobe Photoshop (version 21.2) to make their background brighter.

Combination: Open Shot Video Editor (Version 2.5.1. – libopenshot 0.2.5) was used to create short video (MP4 h.264 with HD 720p 25 fps – 1280x720) by using the materials from the previous two steps. To be more specific, for the familiarization phase a 2-minute video was created, in which every object/subject was named and presented on the screen for 5 seconds. For the main stimuli 59 videos (10 seconds duration/video) were created. In all videos, the sound started with 2 seconds delay to avoid system loading problems during the actual experiment.

2.3 Procedure

The experiment was conducted using PsychoPy (Version 2020.1.3) and consisted of three parts; its total duration was approximately 30 minutes. The experiment starts with the participant sat behind a computer screen and the procedure begins with the picture-naming phase (BNT), followed by the familiarization phase and ends with the Answer-Question (main) phase (Figure 1). Each phase lasted almost 10 minutes. All responses were recorded by a microphone (AKG C1000s). The microphone was placed at 20 cm distance from the participant and with a 5 degree between participant's mouth and microphone.

Figure 1 | Visual Representation of the Experiment Procedure

[3] When the recording had finished, the researcher selected the whole sound and reduced the background noise. Firstly, the creation of the noise profile of the sound was needed, by using the options of Audacity (EFFECT > NOISE REDUCTION > Get Noise Profile). Then, the background noise was reduced by keeping the whole sound selected, (EFFECT > NOISE REDUCTION > Ok).

Picture-naming/BNT phase: The participant was asked to name the pictures that he/she saw in the screen. For details see Appendix B.

Familiarization phase: The participant was watching a short video, in which subjects, objects and action were presented to them. After that, he/she had to watch the same video without sound and name the subjects, objects and action presented to them before. If he/she failed to name at least 23 out of 26 pictures, as they were presented, he/she had to try again.[4] Upon completion of this task, the participant had his/her first trial with the familiarization set of materials (see Appendix D and E).

Answer-Question phase: During that phase, the participants saw one picture and listened to a *who* or a *what* question and he/she had 5 seconds to answer to each question. If the participant accidentally mispronounced a word or cough, he/she could repeat his/her production again. In this task, the participant was instructed to listen carefully and produce a complete sentence with subject, verb and object aloud. All verbal responses were recorded at a rate of 44.1 kHz for offline acoustic analysis.

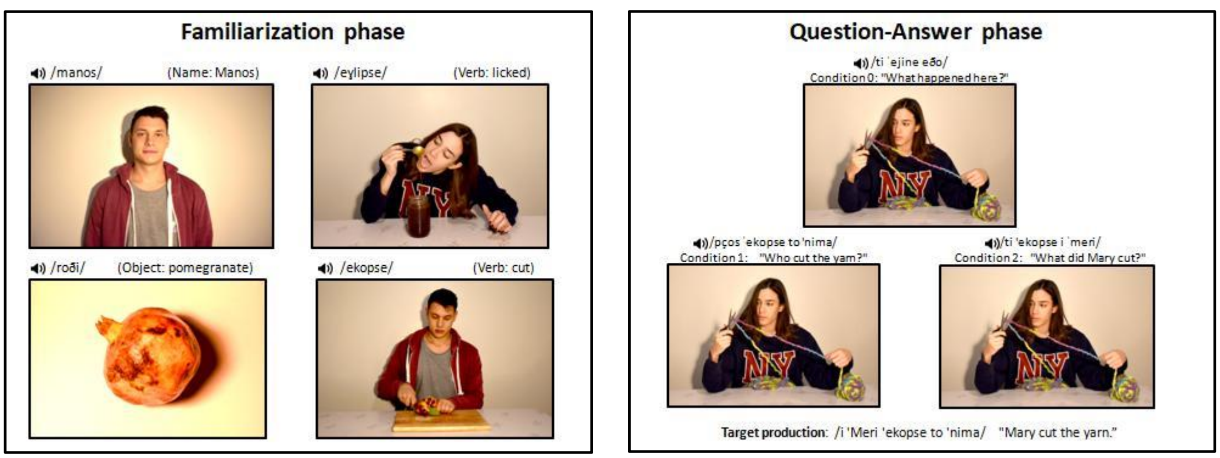


Figure 2 | Examples from the familiarization and main phase of the experiment
Left: Examples from the familiarization phase. Each square represents a screen shot. Right: Examples of the procedure of the Question-Answer phase. Three conditions are presented: Wide focus (upper), Narrow focus on subject (bottom left) and Narrow focus on object (bottom right).

[4] All participants scored 23 or more from the first time. The three missing points were from naming subjects but it was something expected, because early pilot testing showed that participants had poor recall for the names of the people in the pictures. Therefore, in the main phase of the actual experiment, subjects could refer to a sheet which had labeled pictures of the people.

2.3.1. PsychoPy

The PsychoPy software provided two pieces of information for each participant: a) the random sequence of stimuli and b) the number/label of each recording. In total, 3.030 recordings (101 for each participant) were collected. For each participant, the first 47 recordings were from the BNT test and familiarization phase. The remaining 54 recordings were from the main experiment out of which, 18 were the fillers and 36 were the testable productions. Thus, in total, 1080 productions were elicited, in which 2160 target tokens [30 speakers x 12 questions x 3 conditions x 2 focus locations (Subject and Object)] were included. However, during the experiment the PsychoPy software was recording both the recorded question and the participant's reply in each item. Therefore, the first stage of cleaning the data involved manually isolating the participant's productions.

2.4 Acoustic Data Analysis

Of the 2160 speaker tokens, 63 (2.9%) were discarded because (1) the answerer failed to use the correct lexical items; (2) the answerer was disfluent; or (3) the production was poorly recorded. The 2097 remaining tokens were subjected to the acoustic analyses described below.

Consistent with previous work of focus phenomenon on Greek and other languages, as reviewed in Chapter 1, the phonetic realization of prosodic cues is described in terms of the relative changes in the acoustic attributes of duration and F0 at the syllable and word level. Thus, specific acoustic measurements related to duration and F0 were performed at syllable and word level for each of the utterances.

Each utterance was manually annotated at syllable and word level in Praat interface (Boersma & Weenink, 2014). Figure 3 shows a screenshot of the Praat interface with a segmented utterance. Then, using a Praat script (see Appendix F) duration and F0 minimum (F0min), maximum (F0max) and span (F0 span) of each labeled interval was extracted. For each interval, duration was automatic calculated by subtracting the end time from the starting time. Next, the pitch was extracted (with Time step = 0.0 s; Pith floor = 65 Hz; Pitch ceiling = 500) to calculate F0max, F0min and F0span. F0min and F0max values were automatically extracted by built-in Praat commands; F0span was calculated by subtracting F0max from F0min. Finally, these acoustic measures for each participant were automatically stored in a CSV file.

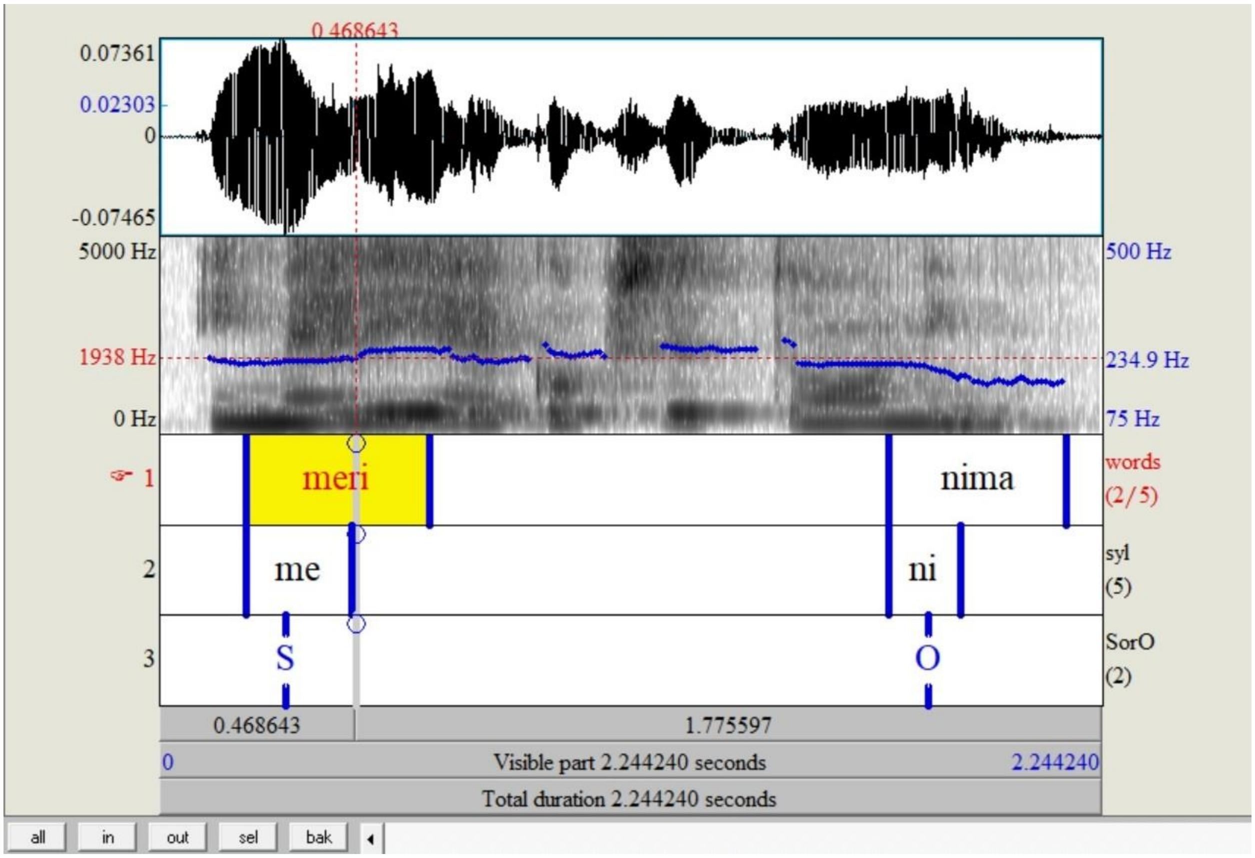


Figure 3 | Screenshot of Praat views of waveform and spectrogram of an utterance with the words and syllable boundaries marked.
The utterance in this example is the following: /i meri ekopse to nima/ (Mary cut the yarn). In the last layer (point tier), S (Subject) corresponds to /meri/ (Mary) and O (Object) corresponds to /nima/ (yarn).

2.5 Descriptive Statistics

The final data file was obtained by merging the two output files from PsychoPy and Praat, and adding additional information from the questionnaire (e.g., Language delay) in R (Version 3.6.2). The variables of interest are the following 8 continuous variables: F0max, F0min, F0span and duration at word (4 variables) and syllable (4 variables) levels.

A descriptive analysis was carried out using the basic functions of R (Version 3.6.2). In other words, descriptive statistics was used for summarizing and visualizing the data, but also, for investigating possible outliers in the variables of interest, before moving to inferential statistics. The experimenter sought only extreme outliers, which showed errors through the recording due to the presence of background noise.

Firstly, the minimum, maximum and the mean value with the standard deviation for each variable of interest were calculated. If there were unexpected values (e.g., zero F0span), the appropriate filters were applied (e.g., F0span > 0) to exclude these cases. Secondly, the researcher created a boxplot and a 5-point summary for each variable of interest. If extreme outliers (i.e., points that beyond an outer fence) were presented, a filter was applied with the criterion of Q3 + 3*IQ for the right fence and Q1 – 3*IQ for the left fence. It should be mentioned that extreme values were excluded for males and females separately. This happened because the values of the variables of interest, namely duration and F0 (max, min and span), vary depending on the participants' gender (Pépiot, 2014).

Finally, the researcher repeated the procedure that she did in the beginning, creating also the corresponding tables and boxplots, which helped her to make some preliminary assumptions for the hypotheses.

2.6 Linear Mixed-Effects Model

Statistical analysis was carried out in R (Version 3.6.2) using the package lme4 (Bates et al., 2015) for the linear mixed models. Linear mixed effects (LME) analysis was performed for each variable of interest. In all the hypotheses test, a α=5% level of significance is used. To assess the significance of differences between levels of in fixed effects, a supplementary lmerTest package, which computes the p-values based on Satterthwaite's approximations (Kuznetsova et al., 2017), was used. It is noteworthy to mention that visual inspection of residual plots for each model was conducted and did not reveal any obvious deviations from homoscedasticity or normality in most of the cases.

A series of mixed effects linear regression models were conducted for word and syllable separately, investigating each phonetic variable of interest F0max, F0min, F0span and duration. The models consist of the following main fixed effects: Gender (Female vs. Male), Group (0: NT vs. 1: High-A vs. 2: Mod-A), Subject/Object (Object vs. Subject) and Condition (0: Wide Focus vs. 1: Focus on Subject vs. 2: Focus on Object), as well as their two-way interactions. Firstly, models with each of the individual fixed effects were compared with a null model with only the random intercepts of subject/participant (30 speakers) and item (8 different words). These random intercepts were included because the same word and syllable is uttered multiple times by the same speaker (Winter, 2019). Log-likelihood ratios were used to evaluate the significance of each of the fixed effects. The Maximum Likelihood criterion was optimized in order to use the ANOVA test for the best fit model selection. Only the effects that were found to be significant were added to each model. An exception in this rule applied in the Group and Subject/Object main effect, as they were the mandatory factors for the research question to be answered. Thus, in Table 3, when these factors were not significant, they were marked with red color. Secondly, this process was repeated with the two-way interactions. For details about the fixed effects of each variable of interest see Table 3 and Appendix G in which there are the formulas and the comparisons with the performance of each model.

Table 3 | Final LME models – Fixed Effects for each variable of interest at word and syllable level

Word Level

		F0max	Duration	F0min	F0span
Main Fixed Effects	Condition	√	√	√	√
	Gender	√	√	√	√
	Subject/Object	√	√	√	√
	Group	√	√	√	√
Two-way interactions	Condition x Subject/Object	√	√	√	√
	Gender x Subject/Object	√	√		√
	Condition x Gender	√			√
	Gender x Group			√	
	Condition x Group	√	√		
	Subject/Object x Group	√	√		

Syllable Level

		F0max	Duration	F0min	F0span
Main Fixed Effects	Condition	√	√	√	√
	Gender	√		√	
	Subject/Object	√	√	√	√
	Group	√	√	√	√
Two-way interactions	Condition x Subject/Object	√	√	√	√
	Gender x Subject/Object			√	√
	Condition x Gender	√			√
	Condition x Group	√		√	
	Subject/Object x Group			√	√

It is important to mention that although the groups differed in BNT score, the inclusion of full-scale BNT did not strengthen the model, so it was excluded. Also, age-related differences would not be expected as all participants had matched based on this criterion, so age was not included in the model.

Chapter 3
Results

This thesis studied the focus realization in Greek speakers, with and without ASD, in acoustic measures of F0max, F0min F0span and duration at syllable and word level. An overview of the data is presented followed by the LME models results, first at word and then at syllable level.

3. 1 Results at word level

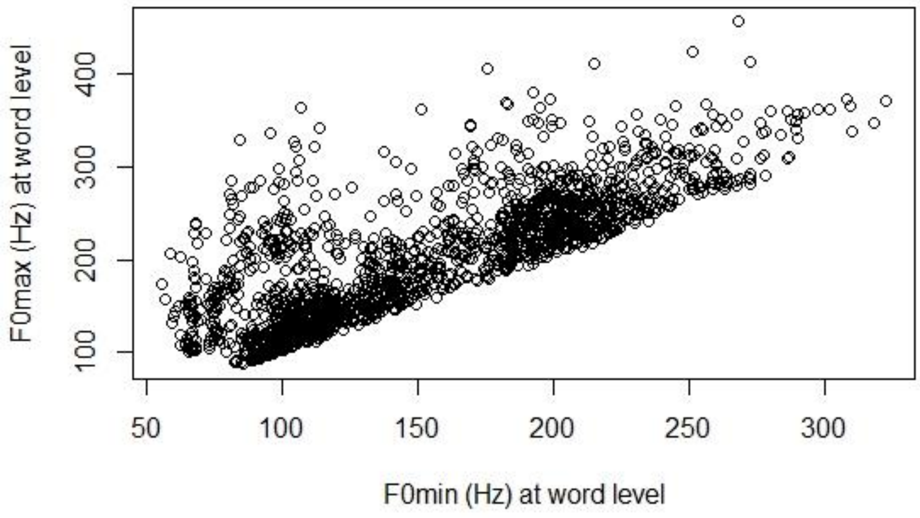
3. 1. 1 Descriptive statistics at word level

Table 4 | Summary of variables of interest at word level

	Mean	SD	min	max	Gender
F0max (Hz)	260.79	42.66	172.93	456.05	female
	151.36	38.39	87.19	320.74	male
F0min (Hz)	203.30	38.79	80.47	322.37	female
	112.45	25.19	55.55	203.72	male
F0span (Hz)	57.49	45.27	3.74	256.74	female
	38.91	34.31	0.00	208.98	male
Duration (sec)	0.38	0.12	0.18	0.86	female
	0.31	0.09	0.13	0.72	male

In the above table (Table 4), the mean value with the standard deviation (SD) and the minimum (min), maximum (max) values of each variable, at *word level*, are presented separately for female and male participants. The production of the target words showed higher values of the F0 variables and longer duration by females, compared to the corresponding productions of males. Moreover, a strong positive correlation (r = .80) between F0min and F0max variables was observed (see Figure 4).

Figure 4 | F0min vs F0max at word level



3. 1. 2 Linear Mixed-Effects model at word level

A summary of the statistical results for the mixed-effects models corresponding to four word level variables is presented in Table 5.[5] The estimations (Est) with the standard errors (SE) and the corresponding t-values with the significance for each of the fixed effects and interactions between them are presented. Also, the variance and the standard deviation (*Std.Dev.)* of each random intercept are reported.

---

[5] It is important to notice that in Table 5, as in Table 7 that it is presented in the next sub-section, only the effects which were significant (at least for one variable) are represented on it. One exception is the Group effect, that for some variables (e.g. F0min and F0span at word level and F0span at syllable level) had no significant differences but as it is the most important factor in this study, all the comparisons between groups are reported. For the rest, in the case that an effect was significant for one variable (e.g. Condition1 for F0max) but not for another (e.g. F0min), the values of it are reported in grey color.

"/" symbol means that this effect or interaction was not included in the model of the corresponding variable.

What is reported in the second column was compared with the base of each model (first column), namely the neutral situation [i.e. Condition: 0 –Wide Focus–, Gender: Females, Subject/Object: Object, Group: 0 –Neurotypical group (NT)–]. For instance, in the fifth row the Condition 1 (FS: Focus on Subject) was compared with the condition 0 (WF: Wide Focus).

Table 5 | Summary of LME at word level

		F0 max (Hz)		Duration (sec)	
		Est. (SE)	**t-values**	**Est. (SE)**	**t-values**
	Intercept	241.15 (10.63)	22.68 ***	0.3667 (0.0216)	17.01 ***
Fixed effects					
Condition WF	Condition FS	-18.68 (3.64)	-5.13 ***	-0.0262 (0.0092)	-2.79 **
	Condition FO	7.70 (3.64)	2.11 *	0.0215 (0.0092)	2.33 *
Females	Males	-106.52 (10.19)	-10.45 ***	-0.0955 (0.0188)	-5.08 ***
Object	Subject	17.89 (3.50)	5.12 ***	-0.0183 (0.0160)	-1.14
NT	**High-A**	23.17 (12.22)	1.90	0.0533 (0.0235)	2.27 *
	Mod-A	30.46 (12.23)	2.49 *	0.0904 (0.0235)	3.84 ***
WF & Object	FS & Subject	10.14 (3.19)	3.18 **	0.0253 (0.0094)	0.27
	FO & Subject	-10.30 (3.18)	-3.24**	-0.0731 (0.0094)	-7.79 ***
Object & NT	Subject & Mod-A	34.11 (12.22)	2.79 **	-0.0134 (0.0094)	-1.43
Females & Object	Males & Subject	-11.85 (2.64)	-4.49 ***	0.0635 (0.0078)	8.15 ***
WF & Females	FS & Males	8.46 (3.23)	2.62 **	/	/
WF & NT	FS & Mod-A	-10.02 (3.89)	-2.58*	-0.0433 (0.0117)	-3.69 ***
	FO & Mod-A	-6.35 (3.89)	-1.64	-0.0273 (0.0117)	-2.73 *
		Variance	***Std.Dev.***	**Variance**	***Std.Dev.***
Random effects	Participant	696.66	26.40	0.0023	0.0481
	Word	3.22	1.80	0.0003	0.0829

		F0 min (Hz)		F0 span (Hz)	
		Est. (SE)	**t-values**	**Est. (SE)**	**t-values**
	Intercept	177.99 (6.84)	26.02 ***	64.82 (8.35)	7.76 ***
Fixed effects					
Condition WF	Condition FS	2.08 (1.84)	1.13	-23.80 (3.40)	-7.00 ***
	Condition FO	-4.54 (1.87)	-2.43 *	9.58 (3.41)	2.81 **
Females	Males	-79.61 (6.55)	-12.16 ***	-27.15 (8.09)	-3.36 **
NT	**High-A**	9.26 (7.47)	1.20	12.42 (9.18)	1.36
	Mod-A	11.38 (7.47)	1.47	15.56 (9.18)	1.70
Object	Subject	39.49 (2.16)	17.82 ***	-20.93 (3.75)	-5.58 ***
WF & Object	FS & Subject	-11.05 (2.60)	-4.26 ***	20.90 (3.67)	5.70 ***
	FO & Subject	4.56 (2.90)	1.76	-15.06 (3.66)	-4.12 ***
Females & Object	Males & Subject	-21.08 (2.15)	-9.80 ***	9.23 (3.04)	3.04 **
WF & Females	FS & Males	/	/	8.70 (3.71)	2.34 *
		Variance	***Std.Dev.***	**Variance**	***Std.Dev.***
Random-effects	Participant	291.65	17.08	404.32	20.11
	Word	0.02	0.16	8.05	2.84

Significance: *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$

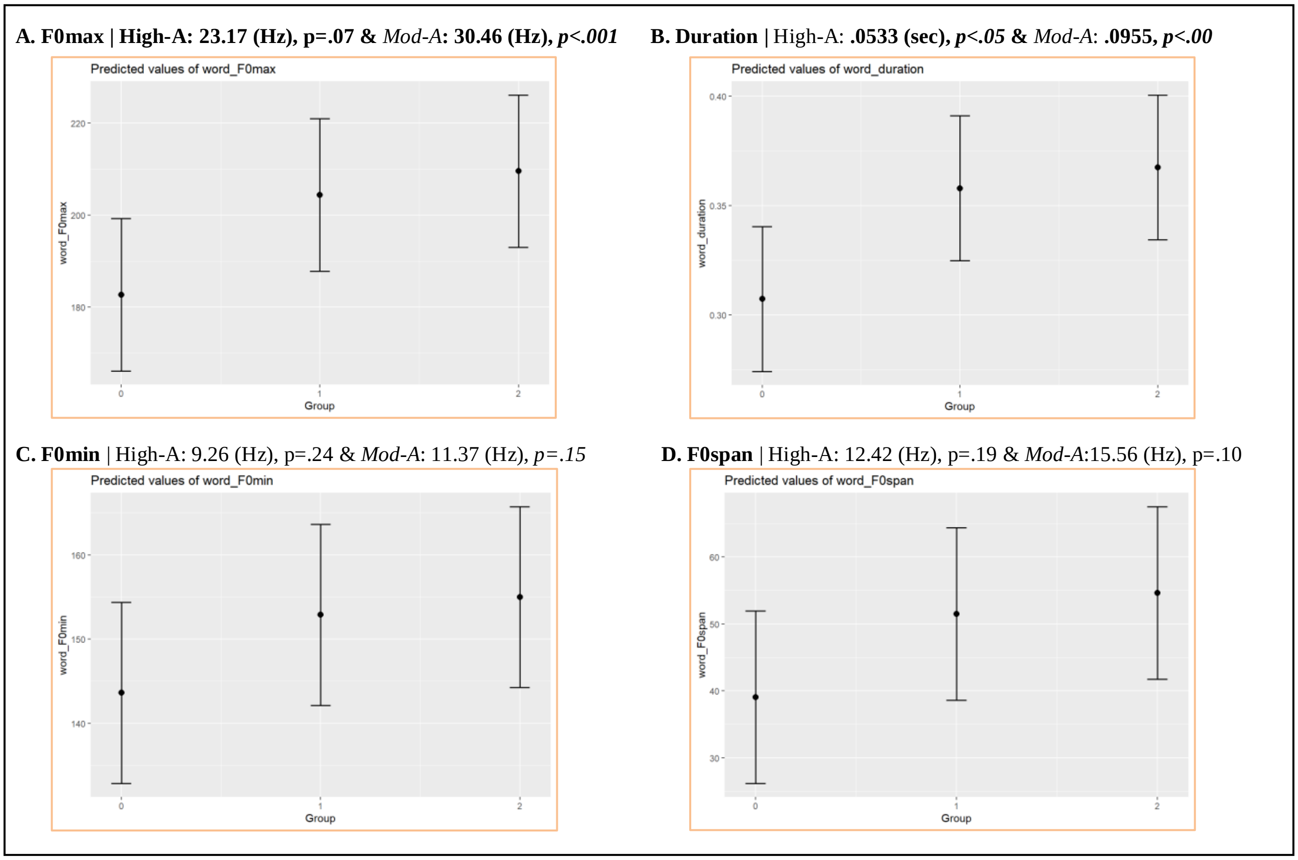
Overall, the results differ in each variable of interest at the word level, but some effects, such as condition, were significant in all of them. First, the two conditions, Focus on Subject (FS) and Focus on Object (FO), differed significantly and in the opposite way from the Wide Focus (WF) condition. These results show that the information structure phenomenon has taken place in these productions. Looking more closely, it is evident that words at FO were produced with higher F0max, longer duration, lower F0min and wider F0span (F0max: $\beta=7.70$, $p<.05$ | duration: $\beta=.0215$, $p<.05$ | F0min: $\beta=-4.54$, $p<.05$ | F0span: $\beta=9.58$, $p<.01$). In contrast, words at FS were produced with lower F0max, shorter duration and narrowed F0span (F0max: $\beta=-18.68$, $p<.001$ | duration: $\beta=-0.0262$, $p<.01$| F0span: $\beta=-23.80$, $p<.001$). However, no significant difference between FS and WF condition was observed in F0min ($\beta=2.08$, $p=.08$). To sum up, the results of the condition effect at word level, it is noticeable that when the target words were produced in the condition that the object was under focus (FO), F0max was higher, duration was longer and F0span was wider than in the other two conditions, namely WF and FS.

Moreover, irrespective of the focus condition that the words were produced, differences between subject and object were observed in all variables, except for duration. Subject was produced with significantly higher F0max ($\beta=17.89$, $p<.001$) and F0min ($\beta=39.49$, $p<.001$), but Object was produced with significantly wider pitch range ($\beta=20.93$, $p<.001$). In duration the production of the subject and the object did not differ significantly ($\beta=-0.02$, $p=.27$) and this implied that there is another factor that can explain the differences between the three focus conditions, namely WF, FS and FO.

The results mentioned above lead the researcher to seek for group differences. When group differences were examined in duration, it was observed that High-A (HFA participants with high language skills) and Mod-A (HFA participants with moderate language skills) demonstrated a longer duration than NT (neurotypical participants), as significant group differences were observed in duration between NT and High-A ($\beta=0.05$, $p<.05$), and Mod-A ($\beta=0.09$, $p<.001$) (see Figure 5B). More specifically, Mod-A exhibited a longer duration of the words when the focus was on subject ($\beta=-0.04$, $p<.001$) and on object ($\beta=-0.03$, $p<.05$) compared to the words that NT produced at wide focus condition. Furthermore, the two-way interaction between the Condition and the Subject/Object effect showed that the subject at FO was produced with significantly shorter duration than the object at WF ($\beta=-0.07$, $p<.001$). No significant difference observed between subject at FS and object at WF ($\beta=0.03$, $p=.79$).

Duration was the only variable in which both groups of participants with HFA differed from NT. Regarding F0max (see Figure 5A), group differences were significant between Mod-A and NT ($\beta=30.46$, $p<.05$), with F0max being higher for the Mod-A group. More specifically, Mod-A demonstrated a significantly higher F0max ($\beta=34.11$, $p<.01$) in subjects when compared with NT in object productions. However, no significant group differences were observed in F0min (High-A: β=9.26, p=.24 | Mod-A: β=11.36, p=.15) and F0span (High-A: β=12.42, p=.19 | Mod-A: β=15.56, p=.10) (see Figure 5C and 5D).

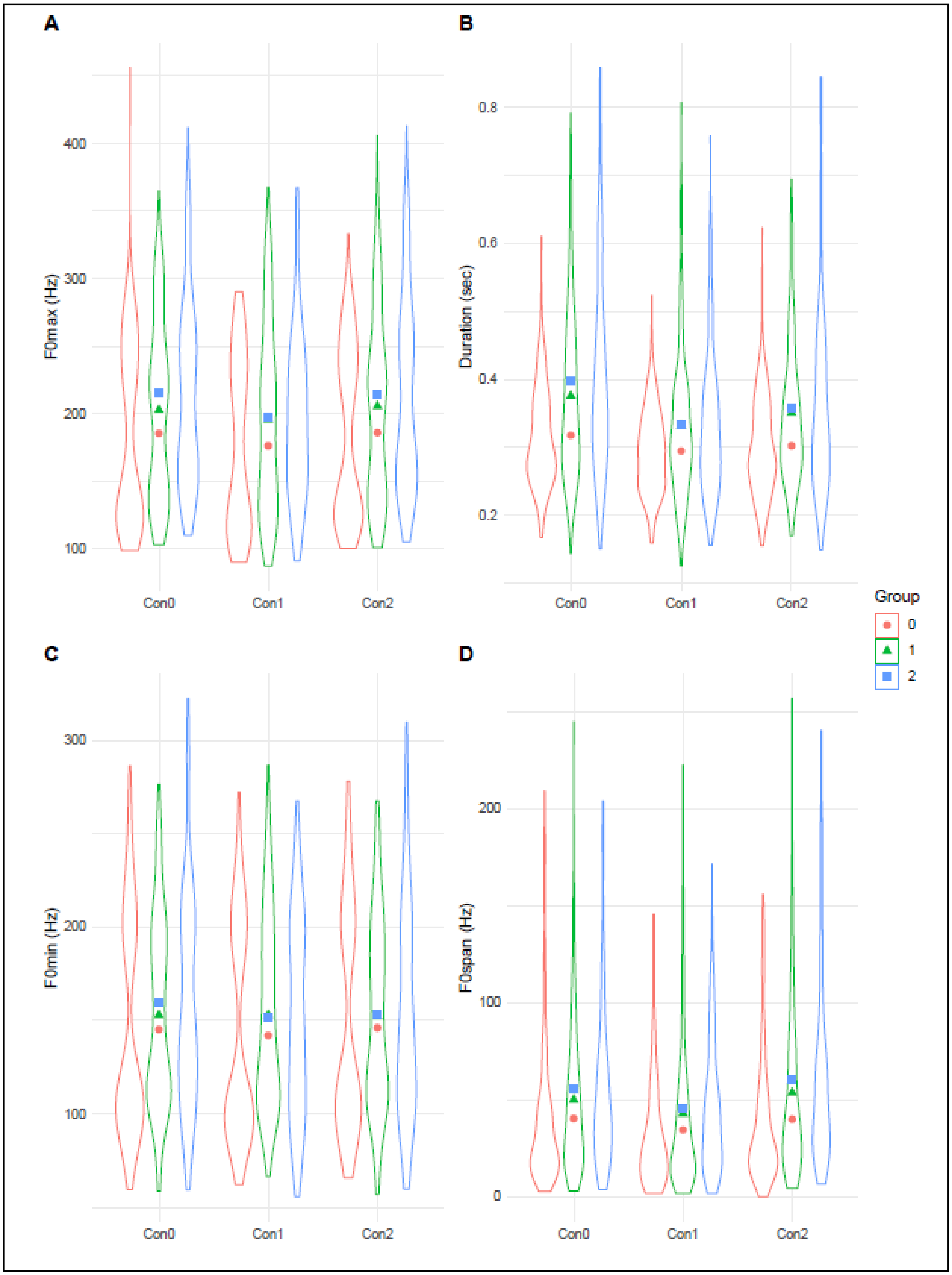
Figure 5 | Main effect of Group (0: NT, 1: High-A, 2: Mod-A) for each of the variables of Interest at Word level



Significantly differences between genders were observed in all the measures of interest (F0max: β=-106.52, p<.001 | duration: β=-.0096, p<.001 | F0min: β=-79.61, p<.0001 | F0span: β=-27.15, p<0.01). More specifically, males had significantly lower F0max and F0min, shorter duration, and narrower pitch range (F0span) than females. In the interaction of the gender and subject/object effect, both genders produced the object with lower F0max and F0min than the subject. Males demonstrated a lower F0max (β=-11.85, p<.001) and F0min (β=-21.08, p<.001) and a longer duration (β=.0064, p<.001) on subject than females on object.

Finally, for a better understanding about the overall behavior of each group in each focus condition, Figure 6 presents a combination of the density plots of each variable (A: F0max, B: duration, C: F0min and D: F0span) by Condition and Group at word level. Mean values of each group are indicated with a red circle for NT, a green triangle for High-A and a blue square for Mod-A. The same colors are used to depict the density plot of each group.

Figure 6 | Variable of interest by Condition and Group at word level



3. 2 Results at syllable level

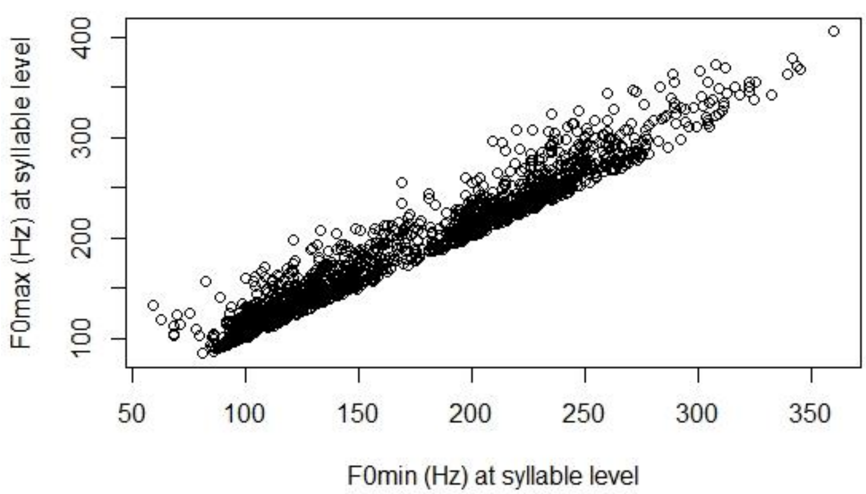
3. 2. 1 Descriptive statistics at syllable level

Table 6 | Summary of variables of interest at syllable level

	Mean	SD	min	Max	Gender
F0max (Hz)	248.34	39.61	105.53	405.77	female
	138.48	29.91	84.65	243.49	male
F0min (Hz)	228.50	31.73	102.04	359.70	female
	122.88	24.30	59.19	227.67	male
F0span (Hz)	19.83	16.52	1.03	88.85	female
	15.60	13.06	0.00	77.41	male
Duration (sec)	0.13	0.04	0.05	0.28	female
	0.13	0.04	0.05	0.33	male

In the above table (Table 6), the mean value with the standard deviation (SD) and the minimum (min), maximum (max) value of each variable, at *syllable level*, are presented separately for female and male participants. It shows that the production of the *stressed (first) syllable* of the target words happened with higher values of the F0 by females, compared to the corresponding productions of males. However, the duration of these syllables does not seem to change between the two genders. Moreover, a very strong positive correlation (r=.97) between F0min and F0max variables was observed (see Figure 7).

Figure 7 | F0min vs F0max at syllable level



3.2.2. Linear Mixed-Effects model at syllable level

A summary of the statistical results for the mixed-effects models corresponding to four syllable level variables is presented in Table 7. As in the Table 5, the estimations (Est), standard errors (SE) with the corresponding t-values for each of the fixed effects and interactions between them are presented. Also, the variance and the standard deviation (*Std.Dev.)* of each of the random intercept are reported.

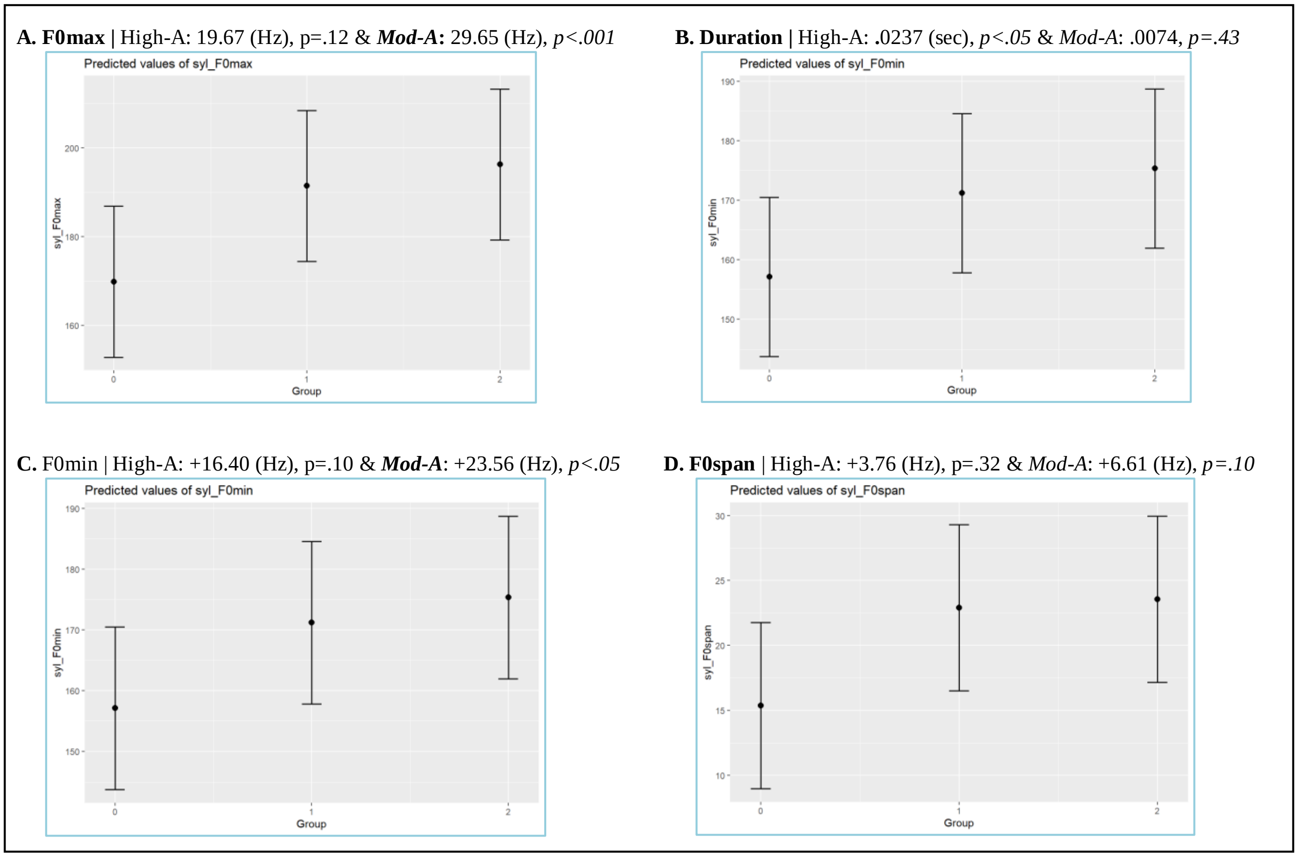
Overall, the results differ in each variable of interest at syllable level, but some effects, such as condition, were significant in all of them. First, the FS condition differed significantly from the WF condition, but FO did not differ from WF. Thus, there was no difference in the first syllable of the words when the object was on focus (FO) and when it was not (FW). This result implies that the marking of the information structure at the syllable level was phonetically detected only in FS condition. Looking more closely to the difference between FS and WF, it is noticeable that the stressed syllables at FS were produced with lower F0max, shorter duration, lower F0min and narrowed F0span (F0max: $\beta=-19.69$, $p<.001$ | duration: $\beta=-.01$, $p<.001$ | F0min: $\beta=-13.57$, $p<.001$ | F0span: $\beta=-6.26$, $p<.001$).

Moreover, irrespective of the focus condition that the stressed syllables were produced, differences between the stressed syllables of subject and object were observed in all variables, except for F0max. Subject was produced with significantly longer duration ($\beta=.03$, $p<.001$), higher F0min ($\beta=8.32$, $p<.001$), but Object was produced with significantly wider pitch range ($\beta=5.74$, $p<.01$). In F0max the production of the stressed syllable of subject and object did not differ significantly ($\beta=2.02$, $p=.27$) and this implies that there is another factor that can explain the differences between the two focus conditions, namely FS and WF.

The results mentioned above lead the researcher to seek for group differences. When group differences were examined in F0max variable, it was observed that Mod-A had higher F0max than NT ($\beta=29.65$, $p<.05$), but no significant group difference was observed between NT and High-A ($\beta=19.67$, $p=.12$) (see Figure 8A). More specifically, Mod-A demonstrated higher F0max of the syllables when the focus was on subject compared to the syllables that NT produced at wide focus condition (WF). Furthermore, the two-way interaction between the Condition and the Subject/Object effect showed that the subject at FS produced with significantly higher F0max than the object at WF ($\beta=21.40$, $p<.001$) and the subject at FO Also, there was no significant difference between subject at FO and object at WF ($\beta=-1.66$, $p=.43$).

On the other hand, High-A demonstrated a longer duration than NT ($\beta=.02$, $p<.05$) (see Figure 8**B**) but no significant differences between these two groups were observed in the remaining values of the syllable level analysis [F0max: β=19.67, p=.12 | F0min: β=16.40, p=.10 | F0span: β=3.76, p=0.33] (see Figure 8A, 8C **&** 8D). However, High-A produced the subject syllable with significantly lower F0min (β=-9.88, $p<.001$) than the object syllable produced by NT. The same behaviour for F0min was also reported for Mod-A in the comparison with NT in the production of subject vs. object syllable, respectively (β=-4.72, $p<.05$). Generally, Mod-A had higher F0max (β=29.65, $p<.05$) and F0min (β=16.39, $p<.05$) than NT.

Figure 8 | Main effect of Group (0: NT, 1: High-A, 2: Mod-A) for each of the variables of Interest at Syllable level



A. F0max | High-A: 19.67 (Hz), p=.12 & ***Mod-A*:** 29.65 (Hz), *p<.001*

B. Duration | High-A: .0237 (sec), *p<.05* & *Mod-A*: .0074, *p=.43*

C. F0min | High-A: +16.40 (Hz), p=.10 & ***Mod-A***: +23.56 (Hz), *p<.05*

D. F0span | High-A: +3.76 (Hz), p=.32 & *Mod-A*: +6.61 (Hz), *p=.10*

Moreover, differences between genders were significant for F0max (β=-111.70, $p<.001$) and F0min (β=-104.16, $p<.001$). In more details, females produced the first syllable of the words with higher F0max and F0min. Gender was not a significant main effect of F0span and duration but the interaction between gender and condition showed that males produced the syllables at FS condition with wider pitch range than the females at WF condition (β=3.61, $p<.01$).

Finally, for a better understanding about the overall behavior of each group in each focus condition, Figure 9 presents a combination of the density plot of each variable (A: F0max, B: duration, C: F0min and D: F0span) by Condition and Group at syllable level. Mean values of each group are indicated with a red circle for NT, a green triangle for High-A and a blue square for Mod-A. The same colors are used to depict the density plots of each group.

Figure 9 | Variable of interest by Condition and Group at syllable level

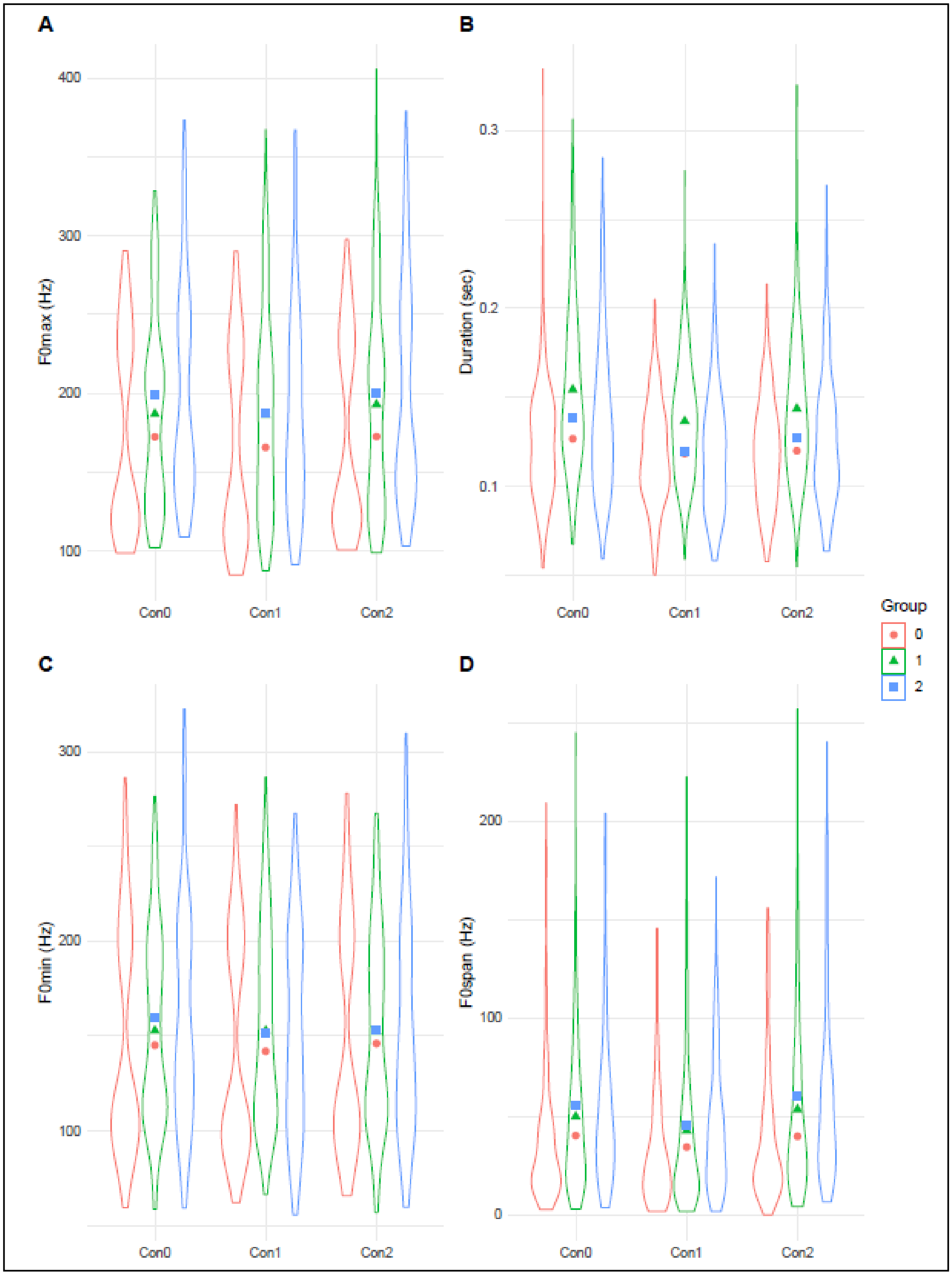


Table 7 | Summary of LME at syllable level

		F0 max (Hz)		Duration (sec)	
		Est. (SE)	**t-values**	**Est. (SE)**	
	Intercept	235.20 (10.73)	21.92 ***	0.1611 (0.0073)	15.98 ***
Fixed-effects					
Condition WF	Condition FS	-19.69 (2.41)	-8.18 ***	-0.0120 (0.0024)	-4.71 ***
Females	Males	-111.70 (10.29)	-10.85 ***	/	/
Object	Subject	2.02 (1.77)	1.15	0.0265 (0.0051)	5.14 ***
NT	**High-A**	19.67 (12.35)	1.59	0.0237 (0.0093)	2.54 *
	Mod-A	29.65 (12.36)	2.40 *	0.0074 (0.0093)	0.79
WF & Object	FS & Subject	21.40 (12.35)	10.16 ***	-0.0071 (0.0035)	-2.11 *
	FO & Subject	-1.66 (2.10)	-0.79	-0.0177 (0.0033)	-5.32 ***
WF & Females	FS & Males	5.31 (2.13)	2.49 *	/	/
WF & NT	FS & Mod-A	-6.56 (2.57)	-2.56 *	/	/
		Variance	***Std.Dev.***	**Variance**	***Std.Dev.***
Random-effects	Participant	746.60	27.32	0.0004	0.0204
	Word	1.74	1.32	0.0000	0.0064
		F0 min (Hz)		**F0 span (Hz)**	
		Est. (SE)		**Est. (SE)**	
	Intercept	215.67 (7.88)	25.53 ***	19.65 (3.40)	5.78 ***
Fixed-effects					
Condition WF	Condition FS	-13.57 (1.82)	-7.43 ***	-6.26 (1.28)	-5.10 ***
Females	Males	-104.16 (7.52)	12.90 ***	/	/
Object	Subject	8.32 (1.89)	4.40 ***	-5.74 (1.67)	-3.44 **
NT	**High-A**	16.39 (9.77)	1.68	3.76 (3.79)	1.00
	Mod-A	23.56 (9.77)	2.41 *	6.61 (3.79)	1.75
WF & Object	FS & Subject	18.15 (1.86)	9.74 ***	3.15 (1.32)	2.38 *
Object & NT	Subject & High-A	-9.88 (1.84)	-5.35 ***	7.37 (1.30)	5.62 ***
	Subject & Mod-A	-4.72 (1.86)	-2.54 *	3.10 (1.32)	2.35 *
Females & Object	Males & Subject	-3.33 (1.54)	-2.16 *	/	/
WF & Females	FS & Males	/	/	3.61 (1.34)	2.70 **
WF & NT	FS & High-A	6.08 (2.26)	2.69 **	/	/
	FS & Mod-A	-5.30 (2.28)	-2.33 *	/	/
		Variance	***Std.Dev.***	**Variance**	***Std.Dev.***
Random-effects	Participant	460.20	21.45	67.23	8.20
	Word	0.00	0.00	1.85	1.36

Significance: *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$

Chapter 4
Discussion

This was a production study that investigated the prosodic encoding of information structure in ASD. This section presents a summary of the findings and the clinical implications of the study. Then, some general results about the information structure interpretation in Greek are presented, followed by the discussion of the limitations of the current study. Finally, areas for future research are proposed.

4.1 ASD & information structure encoding

The main goal of this study was to investigate whether Greek HFA speakers with different language abilities prosodically mark information structure in the same way as neurotypical speakers. A production experiment was conducted, in which information structure of the productions was elicited by manipulating questions that participants had to answer. Three focus conditions were tested, namely, wide focus (WF), focus on subject (FS) and focus on object (FO). Moreover, four acoustic features were measured at word and syllable level, specifically, maximum (F0max) and minimum (F0min) fundamental frequency, pitch range (F0span) and duration of subject and object with their (first) stressed syllable in each condition. HFA participants were matched on age, gender and language skills with neurotypical participants (NT). HFA participants were divided into two groups, the first with high language skills (High-A) and the second with moderate language skills (Mod-A), in order to illuminate the effect of language abilities on this task. To answer the research question of this study a series of linear mixed-effects (LME) model were employed for each phonetic variable.

Overall, it was found that there are differences in the way HFA individuals mark information structure when compared to NT. More specifically, both Mod-A and High-A exhibited longer duration at word level, but only Mod-A demonstrated higher F0max at syllable level, compared to NT. Also, in the majority of the group differences that were observed in the analysis of the measures, the Mod-A group was found to be more often significantly different from the NT group when compared to the High-A group. These results are in line with DePape et al. (2012) findings which demonstrate that the level of the language skills affects the prosodic encoding of the information structure. However, the results of this study are not fully compatible with other aspects of DePape et al. (2012)'s findings and there are methodological reasons that can explain this. First, DePape et al. (2012) tested only males, but as it is evident in this study the gender is a significant factor in the analysis of information structure. Second, DePape et al. (2012) elicited productions in a more structured and walkthrough way, namely by visually indicating the word that was on focus. In contrast, in this study the examiner did not indicate in any way the focused element. Third, they did not test the neutral condition of wide focus (WF), so their comparisons were only between FS and FO conditions and they did not take into account the condition that neither the subject nor the object of the sentence is on focus. In addition to this, they employed analyses of variance (ANOVA), ignoring the repetition of the same words by the same participants which raise the possibility of statistical errors. In contrast in the current study, by including the random intercepts of subject (participant) and item (word) in the LME models, the possibility of this types of errors was reduced. Thus, through the LME models the researcher simultaneously investigated the contribution of multiple elements to acoustic differentiation of the different focus conditions.

Another significant result of this analysis is that group differences were not observed in the range of the F0 neither at word level nor at syllable level. This means that people with HFA use similar pitch range *overall* in the syllable and word productions as NT. This result contradicts the findings of previous studies (Baltaxe et al., 1984; Diehl et al., 2009; Edelson et al., 2007; Nadig & Shaw, 2012), in which HFA participants exhibited wider pitch range than NT. Nevertheless, this lack of F0 range difference can be explained by the findings of Patel et al. (2020). More specifically, they too did not find differences in F0 range of the tested productions, but they found differences in the perceptual ratings of these productions. In other words, their listener-based perceptual data ratings revealed differences in ratings of intonation, in the cases that objective acoustic measures did not. Thus, they suggested that there are differences between participants with ASD and NT, as the listeners perceived these intonation differences, but they may not be due exclusively to differences in the F0 measurements examined if they are not companied with the times of pitch rises and falls into these productions. This point will be analyzed more at the general findings of this study.

Botinis & Bannert (2003) detected a relationship between gender and information structure in neurotypical participants. The findings of this study are compatible with this, as differences in the two-way interaction of gender and the focus condition were observed. This implies that this difference is also present at the participants with ASD. Thus, studying more the gender differences in information structure phenomenon can bring in new evidence and expand our knowledge.

4.2 Clinical Implications

The results suggest that at least some of the heterogeneity of prosodic use among HFA participants is related to the level of language abilities. Regardless of subgroup differences, because prosodic cues are mainly processed unconsciously by typical listeners, inappropriate use of prosody may be interpreted at conscious level by listeners (DePape et al., 2012; Patel et al., 2020). Such speakers will be negatively deemed as less engaged in communication (R. Paul et al., 2005; Shriberg et al., 2001), which could make it more difficult for them to have a typical everyday communication with their family members, friends and colleagues. The speech pathologists need to be aware and differentiate among the types of prosody-voice involvements in an individual with ASD, remaining alert for language skills involvements. It is, therefore, important that the therapists first understand the details of prosodic use in different HFA subgroups and then apply any remediation strategy.

4.3 General findings of information structure encoding

Let us now turn to other issues that involve the validity of this study and the extent to which these results can be generalized in the prosodic encoding of information structure. First and most important, the variables of interest (F0max, F0min, F0span and duration) presented different behavior at word and syllable level. More specifically, condition differences were found in all variables in both levels (word and syllable), with syllable level showing that the difference was only between wide focus (WF) and focus on subject (FS) condition. At word level, variables differed significantly in all the comparisons of the wide focus (WF) condition with the other two (FS and FO: Focus on object). Thus, it is evident that information structure phenomenon had taken place in all these measures at word level. These results lead to the implication that the focus phenomenon goes beyond the borders of the stressed syllable of words for both HFA and NT individuals.

In addition to the differences between conditions, an interesting finding was that in the majority of the measures, the subject was more phonetically prominent than the object for both HFA and NT individuals. This result implies that the difference between focus conditions could not be specified by comparing solely the subject with the object element of a sentence. Furthermore, this result raises the question of whether the word order could play a role in Greek information structure analysis, as the word order in Greek is quite free, with VSO and other orders as frequent alternatives of SVO, which is the predominant word order (Holton et al., 2016). Gryllia (2008) explored the effect of the sentence location in the focused object but she did not find any important relationship between the phonetic realization of the focused object and its location in the sentence. Thus, probably the syntactic structure that the participants chose to use sheds more light on information structure research than does the location of the object in the sentence. In other words, it is not only the sentence location of the target focused element that should be studied but generally the syntactic domain in which the focused element is produced.

4.4 Limitations and Future Research

The present study investigated the prosodic encoding of information structure in HFA adolescent and adults within a specific context. For future research, it would be helpful to systematically assess prosody in less structured tasks of HFA individuals and measure all the elements on sentence or discourse. Although the current study makes a significant contribution to the literature in terms of detecting differences in prosody of Greek HFA individuals related to their language abilities, future work should aim to replicate these results in a larger sample, whose utterances will be segmented by independent transcribers.[6] More importantly, the present study identified longer duration of the words in the HFA groups and this finding appeared to be driven by the HFA who had moderate language skills.[7] Additional evidence is needed to support this conclusion and would have broader implications for speech-language interventions tailored for HFA population based on their language abilities. Furthermore, the present study focused solely on prosodic production. However, further insight into the prosodic profile in HFA individual may be gained through investigation of receptive prosody skills.

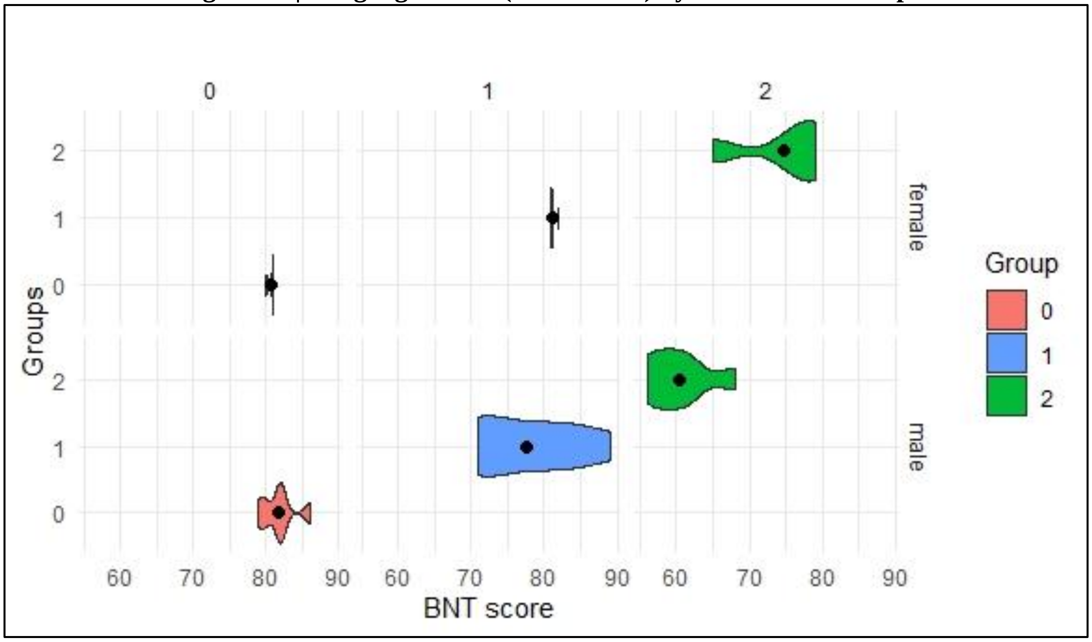
Further research on the relationship between the gender and language abilities of HFA individuals might extend the knowledge of their prosodic profile differences. In the current study, HFA females were observed to have better language skills than HFA males (see Figure 10). As far as I am aware, there is no investigation of this claim in the existing literature. The only evidence there is in literature reports that males and females have different clinical presentation with males having greater externalising and social problems than females (Mandy et al., 2012) and less motivation for social contact (Sedgewick et al., 2016). These findings, together with the results of this study generate the question whether specific aspects of language

[6] It is important to note, that the present results rely solely on experimenter's segmentation. Higher reliability will be gained, only if speech is segmented by transcribers who were blind to the participants' diagnosis.

[7] A post analysis that was conducted in a subgroup of participants gave the same pattern of results, but demonstrated the need for more precise segmentation in the future.

differences, such as the appropriate prosody, could be related to the reduced social skills, especially in males with ASD.

Figure 10 | Language scores (scale: 60-90) by Gender and Group



Last but not least, the present study made several contributions to literature in terms of analyzing phonetically the linguistic expression of information structure, as it showed that in future studies additional acoustic measures of F0 need to be examined for specifying the relative times of pitch rises and falls. Thus, this study is the starting point in a mixed phonetically-phonologically approach in the research of the prosodic encoding of information structure. So far, most of the researchers in Greek followed the AM framework for studying intonation, in which they used labels of pitch accents. As Arvaniti states research should look beyond the AM (Autosegmental Metrical model of intonational phonology) because "the relationship between intonation and F0 is not straightforward" (Arvaniti, 2019, p. 1). Thus, the results of this study indicate the necessity of additional cues, such as the tonal alignment from the phonological framework, in order to reach more sufficient results in studying information structure phenomenon in future samples.[8]

[8] Tonal alignment refers to the temporal implementation of fundamental frequency (F0) movements with respect to the segmental string.

Conclusion

This thesis was the first attempt to investigate the acoustic correlates of Greek-speaking individuals with ASD. The motivation for the study stemmed from the emerging tensions in literature regarding how HFA demonstrate the prosodic encoding of information structure and the fact that there is no published work that tackles this topic in the Greek literature.

The current study used rigorous scientific methods to explore several important questions concerning the acoustic correlates of information structure in ASD. More specifically, it was a production study which explored whether and if so, how Greek speakers with HFA and different levels of language abilities mark prosodically the information structure to encode certain focus conditions, in comparison to neurotypical (NT) speakers. It was found that there are differences in the way that HFA mark information structure compared to NT. More specifically, both Mod-A and High-A exhibited longer duration at word level. Also, in the majority of the group differences that were observed, Mod-A group was found to be more often significantly different from the NT group than the High-A. Gender differences were also investigated in this study, but future work should aim to replicate these results in a larger sample.

Lastly, the results of this thesis provide practical implications for speech pathologists who work with people with ASD and guide then to improve their communication skills. The findings of this thesis also provide a contribution to the understanding of how the level of language abilities relates to the prosodic encoding of information structure. While the limitations of this study, such as the small sample of participants and the necessity of additional measures expect from purely phonetic acoustic cues, need to be acknowledged, the study can be considered as the starting point of prosodic encoding information structure in Greek-speaking individuals with ASD.

References

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (5th ed.)*. https://doi.org/10.1176/appi.books.9780890425596

Arvaniti, A. (2007). Greek PhoneticsThe State of the Art. *Journal of Greek Linguistics*, *8*(1), 97–208. https://doi-org.ezproxy.leidenuniv.nl:2443/10.1075/jgl.8.08arv

Arvaniti, A. (2019). Crosslinguistic variation, phonetic variability, and the formation of categories in intonation. *Proceedings of ICPhS 2019*. https://assta.org/proceedings/ICPhS2019/

Arvaniti, A., & Baltazani, M. (2005). Intonational analysis and prosodic annotation of Greek spoken corpora. In *Prosodic typology: The phonology of intonation and phrasing.* (Vol. 84, pp. 117–169). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199249633.001.0001

Arvaniti, A., & Baltazani, M. (2014). *GRToBI*. https://amaliaarvaniti.wixsite.com/grtobi

Arvaniti, A., Ladd, D. R., & Mennen, I. (2006). Tonal Association and Tonal Alignment: Evidence from Greek Polar Questions and Contrastive Statements. *Language & Speech*, *49*(4), 421–450. https://doi.org/10.1177/00238309060490040101

Atsidakou, M., Siganou, Z., & Skoulakari, E. (2011). *Reliability and Validity of Short Edition of Boston Diagnostic Assessment Examination and Boston Naming Test for Aphasia in Greek* [Bachelor’s thesis, TEI of Western Greece]. Patras.

Baltaxe, C., Simmons, J. Q., & Zee, E. (1984). Intonation patterns in normal, autistic and aphasic children. *10th Internation Congress of Phonetic Sciences*, 713–718.

Baltazani, M. (2003). *Broad Focus Across Sentence Types in Greek*. Eighth European Conference on Speech Communication and Technology, Geneva.

Baltazani, M., & Jun, S. A. (1999). Focus and topic intonation in Greek. *14th International Congress of Phonetic Sciences (ICPhS)*, 1305–1308.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Boersma, P., & Weenink, D. (2014). *Praat: Doing phonetics by computer (version 6.1.21)*. http://www.praat.org/

Bosch, P., & Sandt, R. van der. (1999). *Focus: Linguistic, Cognitive, and Computational Perspectives*. Cambridge University Press.

Botinis, A., & Bannert, R. (2003). Focus and gender interactions and prosodic correlates. *PHONUM 9*, 105–108.

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

DePape, A. M. R., Chen, A., Hall, G. B., & Trainor, L. J. (2012). Use of Prosody and Information Structure in High Functioning Adults with Autism in Relation to Language Ability. *Frontiers in Psychology*, *3*(72). https://doi.org/10.3389/fpsyg.2012.00072

Diehl, J. J., Watson, D., Bennetto, L., Mcdonough, J., & Gunlogson, C. (2009). An acoustic analysis of prosody in high-functioning autism. *Applied Psycholinguistics*, *30*(3), 385–404. https://doi.org/10.1017/S0142716409090201

Dilley, L. C. (2005). *The phonetics and phonology of tonal systems* [Doctoral dissertation, Massachusetts Institute of Technology]. https://dspace.mit.edu/handle/1721.1/30274

Edelson, L., Grossman, R., & Tager-Flusberg, H. (2007). *Emotional prosody in children and adolescents with autism*. Poster session presented at the annual international meeting for Autism Research, Seattle, WA.

Fosnot, S. M., & Jun, S.-A. (1999). *Prosodic characteristics in children with stuttering or autism during reading and imitation.* 1925–1928.

Fusaroli, R., Lambrechts, A., Bang, D., Bowler, D. M., & Gaigg, S. B. (2017). "Is voice a marker for Autism spectrum disorder? A systematic review and meta-analysis." *Autism Research*, *10*(3), 384–407. https://doi.org/10.1002/aur.1678

Goldsmith, J. A., Riggle, J., & Alan, C. L. (1995). *The handbook of phonological theory*.

Gryllia, S. (2008). *On the nature of preverbal focus in Greek: A theoretical and experimental approach: proefschrift*. LOT.

Gussenhoven, C. (1999). On the limits of focus projection in English. In *Focus: Linguistic, Cognitive, and Computational Perspectives* (pp. 43–55). Cambridge University Press.

Holton, D., Mackridge, P., & Philippaki-Warburton, I. (2016). *Greek: An Essential Grammar of the Modern Language* (2nd edition). Routledge.

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

Krüger, M., Cangemi, F., Vogeley, K., & Grice, M. (2018). Prosodic Marking of Information Status in Adults with Autism Spectrum Disorders. *9th International Conference on Speech Prosody 2018*, 182–186. https://doi.org/10.21437/SpeechProsody.2018-37

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, *82*(1), 1–26. https://doi.org/10.18637/jss.v082.i13

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

Lohfink, G., Katsika, A., & Arvaniti, A. (2019). *Variability and category overlap in the realization of intonation*. 701–705. https://repository.ubn.ru.nl/handle/2066/220393

Mandy, W., Chilvers, R., Chowdhury, U., Salter, G., Seigal, A., & Skuse, D. (2012). Sex Differences in Autism Spectrum Disorder: Evidence from a Large Sample of Children and Adolescents. *Journal of Autism and Developmental Disorders*, *42*(7), 1304–1313. https://doi.org/10.1007/s10803-011-1356-0

Marinis, T., Terzi, A., Kotsopoulou, A., & Francis, K. (2013). Pragmatic abilities of high-functioning Greek-speaking children with autism. *Psychology*, *20*(3), 321–337.

McCann, J., & Peppé, S. (2003). Prosody in autism spectrum disorders: A critical review. *International Journal of Language & Communication Disorders*, *38*(4), 325–350. https://doi.org/10.1080/1368282031000154204

Nadig, A., & Shaw, H. (2012). Acoustic and Perceptual Measurement of Expressive Prosody in High-Functioning Autism: Increased Pitch Range and What it Means to Listeners. *Journal of Autism and Developmental Disorders*, *42*(4), 499–511. https://doi.org/10.1007/s10803-011-1264-3

Papakyritsis, I., Kastani, I., & Nerantzini, M. (2019). Phonological awareness and sonority in Greek children: Developmental data and clinical perspectives. *Clinical Linguistics & Phonetics*, *34*(8), 756–773. https://doi.org/10.1080/02699206.2019.1697371

Patel, S. P., Nayar, K., Martin, G. E., Franich, K., Crawford, S., Diehl, J. J., & Losh, M. (2020). An Acoustic Characterization of Prosodic Differences in Autism Spectrum Disorder and First-Degree Relatives. *Journal of Autism and Developmental Disorders*, *50*(8), 3032–3045. https://doi.org/10.1007/s10803-020-04392-9

Paul, R., Augustyn, A., Klin, A., & Volkmar, F. R. (2005). Perception and Production of Prosody by Speakers with Autism Spectrum Disorders. *Journal of Autism and Developmental Disorders*, *35*(2), 205–220. https://doi.org/10.1007/s10803-004-1999-1

Paul, Rhea, Bianchi, N., Augustyn, A., Klin, A., & Volkmar, F. R. (2008). Production of syllable stress in speakers with autism spectrum disorders. *Research in Autism Spectrum Disorders*, *2*(1), 110–124. https://doi.org/10.1016/j.rasd.2007.04.001

Pépiot, E. (2014). Male and female speech: A study of mean f0, f0 range, phonation type and speech rate in Parisian French and American English speakers. *7th International Conference on Speech Prosody 2014*, 305–309. https://doi.org/10.21437/SpeechProsody.2014-48

Rao, P., Sanghvi, N., Mixdorff, H., & Sabu, K. (2017). Acoustic correlates of focus in Marathi: Production and perception. *Journal of Phonetics*, *65*, 110–125. https://doi.org/10.1016/j.wocn.2017.07.002

Sedgewick, F., Hill, V., Yates, R., Pickering, L., & Pellicano, E. (2016). Gender Differences in the Social Motivation and Friendship Experiences of Autistic and Non-autistic Adolescents. *Journal of Autism and Developmental Disorders*, *46*(4), 1297–1306. https://doi.org/10.1007/s10803-015-2669-1

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

Shriberg, L. D., Rhea, P., McSweeny, J. L., Ami, K., Cohen, D. J., & Volkmar, F. R. (2001). Speech and Prosody Characteristics of Adolescents and Adults With High-Functioning Autism and Asperger Syndrome. *Journal of Speech, Language, and Hearing Research*, *44*(5), 1097–1115. https://doi.org/10.1044/1092-4388(2001/087)

Simos, P. G., Kasselimis, D., & Mouzaki, A. (2011). Age, gender, and education effects on vocabulary measures in Greek. *Aphasiology*, *25*(4), 475–491. https://doi.org/10.1080/02687038.2010.512118

Simos, P. G., Sideridis, G. D., Protopapas, A., & Mouzaki, A. (2011). Psychometric Evaluation of a Receptive Vocabulary Test for Greek Elementary Students. *Assessment for Effective Intervention*, *37*(1), 34–49. https://doi.org/10.1177/1534508411413254

Skopeteas, S. (2016). Information structure in Modern Greek. In C. Féry & S. Ishihara (Eds.), *Handbook of Information Structure*. Oxford University Press.

Spreen, O., & Risser, A. H. (2003). *Assessment of Aphasia*. Oxford University Press, USA.

Terzi, A., Marinis, T., & Francis, K. (2016). The Interface of Syntax with Pragmatics and Prosody in Children with Autism Spectrum Disorders. *Journal of Autism and Developmental Disorders*, *46*(8), 2692–2706. https://doi.org/10.1007/s10803-016-2811-8

Terzi, A., Marinis, T., Kotsopoulou, A., & Francis, K. (2014). Grammatical Abilities of Greek-Speaking Children with Autism. *Language Acquisition*, *21*(1), 4–44. https://doi.org/10.1080/10489223.2013.855216

Tripolitou, M., & Chaida, A. (2011). *High functioning autism and prosody of sentence types in Greek: A case study* (A. Botinis, Ed.; pp. 143–146). International Speech Communication Association.

Winter, B. (2019). *Statistics for Linguists: An Introduction Using R*. Routledge.

World Health Organization. (2019). *International statistical classification of diseases and related health problems (11th ed.)* (1st Edition). https://icd.who.int/

Appendices

Appendix A: Information & Consent forms

Information for participants of scientific research

Leiden University Centre for Linguistics

Universiteit Leiden

Supervisor: Prof. dr. Yiya Chen
Experimenter: Dafni Vaia Bagioka

Study title: Experiment on Greek Phonology and Phonetics

Dear participant,

We would like to ask you to take part in a study in which we hope to gain more understanding about the Greek language. We aim to do this by a phonetic experiment.

Procedure of the experiment

The experiment consists of two sessions and the total duration of it will be approximately 30 minutes. The procedure starts from Session A and ends to Session B.

Session A: Picture-naming (10 minutes): You will see pictures and you will be asked to name them.

Session B: Answer-Question Game (20 minutes): You will see one picture and you will listen to one question. You will have to answer to each question. Before the experiment begins, there will be a familiarization phase with the pictures.

During the whole procedure, there will be a microphone to record your answers.
Please speak in a natural manner, not too fast and not too slow. It is the best to speak as if you were actually talking to another person (e.g. on the phone). There is no need to worry about making mistakes, as there is no correct answer. If you accidentally mispronounce a word or cough, you would have the chance to repeat your response.

Voluntary participation

Participation in this study is voluntary and you can decide to withdraw at any time, without providing a reason.

Personal Data

All information collected with regard to this study will be treated strictly confidentially. All data will be processed and stored anonymously and they data will not be accessible to unauthorized people and we will not allow individual participants to be personally identified.

1/2

Information for participants of scientific research

Coronavirus (COVID-19)
We would like to assure you that we will strictly apply all the measures proposed by the National Agency for Public Health, both for the room and for the equipment that will be used to conduct the experiment and masks and disposable gloves will be offered free of charge to the participants.

This research is coordinated by the master student Bagioka Dafni Vaia. Please contact her via a phone call or an email message if you have any questions or comments about this study. You can find her contact information in the end of this form.

Complaints

Should you find that you have been incorrectly or insufficiently informed about participation in this study? If you have any complaints about the way this study was performed or the way you have been treated as a participant, it is recommended that you discuss this with the experimenter of the study. If you do not wish to do so, or in case that does not resolve the issue, you can also lodge a complaint with the Leiden University Centre for Linguistics (LUCL). Please find LUCL's contact information in the end of this form.

Consent
In order for you to participate in this study, we require your consent. You can confirm your willingness to participate using the attached consent form.

Contact information

Supervisor: Yiya Chen
Telephone: [redacted]
E-mail: [redacted]

Experimenter: Dafni Vaia Bagioka (Dafni)
Telephone: [redacted] (Greece)/ [redacted] (the Netherlands)
E-mail: [redacted]

Leiden University Centre for Linguistics (LUCL):
Address: [redacted], Leiden, The Netherlands
Telephone: [redacted]
E-mail: [redacted]

Consent Form

Leiden University Centre for Linguistics

Universiteit Leiden

Supervisor: Prof. dr. Yiya Chen
Experimenter: Dafni Vaia Bagioka

Study title: Experiment on Greek Phonology and Phonetics

Informed Consent

By signing this form, you confirm that you have read and understood the participant information form. By signing this form, you also confirm that you agree to the study procedure described in the participant information form.

I have read and understood the participant information form and I agree to participate in this study.

Date: ………………………….. Place: …………………………….

Name: …………………………… Signature: ………………………

Information for participants' parent/caregiver of scientific research

Leiden University Centre for Linguistics

Universiteit Leiden

Supervisor: Prof. dr. Yiya Chen
Experimenter: Dafni Vaia Bagioka

Study title: Experiment on Greek Phonology and Phonetics

Dear parent/caregiver,

We would like to ask you your permission for conducting a study in which we hope to gain more understanding about the Greek language. We aim to do this by a phonetic experiment.

Procedure of the experiment

The experiment consists of two sessions and the total duration of it will be approximately 30 minutes. The procedure starts from Session A and ends to Session B.

Session A: Picture-naming (10 minutes): The participant will see pictures and he/she will be asked to name them.

Session B: Answer-Question Game (20 minutes): The participant will see one picture and he/she will listen to one question. He/she will have to answer to each question.

Before the experiment begins, there will be a familiarization phase with the pictures.

During the procedure, there will be a microphone to record participant's answers. Participants should speak in a natural manner, not too fast and not too slow. It is the best to speak as if he/she was actually talking to another person (e.g. on the phone). There is no need to worry about making mistakes, as there is no correct answer. If he/she accidentally mispronounce a word or cough, he/she would have the chance to repeat his/he response.

Voluntary participation

Participation in this study is voluntary and the participant can decide to withdraw at any time, without providing a reason.

Personal Data

All information collected with regard to this study will be treated strictly confidentially. All data will be processed and stored anonymously and they data will not be accessible to unauthorized people and we will not allow individual participants to be personally identified.

Information for participants' parent/caregiver of scientific research

Coronavirus (COVID-19)
We would like to assure you that we will strictly apply all the measures proposed by the National Agency for Public Health, both for the room and for the equipment that will be used to conduct the experiment and masks and disposable gloves will be offered free of charge to the participants.

This research is coordinated by the master student Bagioka Dafni Vaia. Please contact her via a phone call or an email message if you have any questions or comments about this study. You can find her contact information in the end of this form.

Complaints

Should you find that you have been incorrectly or insufficiently informed about participation in this study? if you have any complaints about the way this study was performed or the way you have been treated as a participant, it is recommended that you discuss this with the experimenter of the study. If you do not wish to do so, or in case that does not resolve the issue, you can also lodge a complaint with the Leiden University Centre for Linguistics (LUCL). Please find LUCL's contact information in the end of this form.

Consent
In order for your adolescent to participate in this study, we require your consent. You can confirm your willingness for your adolescent to participate using the attached consent form for participants' parent/caregiver.

Contact information
Supervisor: Yiya Chen
Telephone: [redacted]
E-mail: [redacted]

Experimenter: Dafni Vaia Bagioka (Dafni)
Telephone: [redacted] (Greece)/ [redacted] (the Netherlands)
E-mail: [redacted]

Leiden University Centre for Linguistics (LUCL):
Address: [redacted] Leiden, The Netherlands
Telephone: [redacted]
E-mail: [redacted]

Consent Form for participants' parent/caregiver

Leiden University Centre for Linguistics

Supervisor: Prof. dr. Yiya Chen
Experimenter: Dafni Vaia Bagioka

Study title: Experiment on Greek Phonology and Phonetics

Informed Consent

By signing this form, you confirm that you have read and understood the participant information form for participants' parent/caregiver. By signing this form, you also confirm that you agree to the study procedure described in the participant information form for participants' parent/caregiver.

I have read and understood the participant information form and I agree my adolescence to participate in this study.

Date: …………………………..   Place: ……………………………

Name: ……………………………   Signature: ………………………

Appendix B: Boston Naming Test (BNT)

BNT is a neuropsychological assessment tool for measuring confrontational word retrieval. The English BNT has been published in two versions: the standard version, consisting of 60 items and the short one, consisting of 15 items (Spreen & Risser, 2003). BNT has been translated and standardized in Greek, in both versions (Simos, Kasselimis, et al., 2011; Simos, Sideridis, et al., 2011). The standard version consists of 45 items and the short version of 20 items.

This study used the standard version. The examiner begins with Item 1 and continues through Item 45. In particular, the participant is shown target stimuli and is asked to identify each target item within a 20-second interval. If the patient fails to give the correct response, the examiner provides a phonemic cue ("The word starts with /initial phoneme/"). If the participant is still unable to name the object correctly or blatantly misconstrues the image, a semantic cue is given as administration criteria are provided in the Test Manual. After the patient completes the test, the examiner scores with 2 points correct response without help, with one point responses with help (phonemic or/and semantic) and with 0 points wrong or no response (Atsidakou et al., 2011). Participants' responses were recorded automatically in PsychoPy 3 data folder.

Appendix C: Background Questionnaire

Form for Therapists

Please answer the following questions as accurately as possible. All information is for statistical purpose only, and will be kept strictly confidential.

Information of Therapist: (Optional, in case some clarification is needed):

Full name:___________________________ Contact number:____________________

Place: ______________________________ Telephone number: __________________

Information of Participant:

Initials: _____ Date of Birth: __________ Gender: M / F

Does he/she have a diagnosis of hearing problem or/and vision difficulties? Yes/ No

Has he/she been diagnosed with Autism Spectrum Disorder (ASD)? Yes / No

Has a family member been diagnosed with ASD? Yes / No

Did he/she experience early language delay? Yes / No

Is Greek his/her native language? Yes / No

If Greek is not his/her native language, which other language(s) does he/she speak and from what age?

____
____
____

Form for participants

Please answer the following questions as accurately as possible. All information is for statistical purpose only, and will be kept strictly confidential.

Initials: _____ Date of Birth: __________ Gender: M / F

Do you have a diagnosis of hearing problem or/and vision difficulties? Yes/ No

Have you been diagnosed with Autism Spectrum Disorder (ASD)? Yes / No

Has a family member been diagnosed with ASD? Yes / No

Have you experience early language delay? Yes / No

Is Greek your native language? Yes / No

If Greek is not your native language, which other language(s) do you speak and from what age?

____
____
____

Appendix D: Stimuli of familiarization and production experiment

Full items are recoverable as follows: Condition 0 always asks ‘‘What happened here?’’. Questions 1 and 2 are wh-questions about the subject and object, respectively.

Set ID	IPA transcription	Gloss
Fam[9]	0. [ti ˈejine eðo]	A. What happened here?
	1. [pços ˈerikse to ˈlaði]	B. **Who** dropped the (jug of) oil?
	2. [ti ˈerikse o ˈmanos]	C. **What** did Manos drop?
	Target: [o ˈmanos ˈerikse to ˈlaði]	Target: Manos dropped the oil.
A	[o ˈmanos ˈefaje to ˈmilo]	Manos ate the apple.
B	[i ˈneli ˈepline to ˈmilo]	Neli washed the apple.
C	[i ˈmeri ˈekopse to ˈnima]	Mary cut the yarn.
D	[i ˈmeri ˈeblekse to ˈnima]	Mary intertwined the yarn.
E	[o ˈmanos ˈefaje to ˈmeli]	Manos ate the honey.
F	[o ˈmanos ˈerikse to ˈmeli]	Manos dropped the honey.
G	[i ˈmeri ˈeɣlipse to ˈmeli]	Mary licked the honey.
H	[i ˈneli ˈevapse ti ˈlima]	Neli painted the nail file.
I	[i ˈmeri ˈespase ti ˈlima]	Mary broke the nail file.
J	[i ˈmeri ˈefaje to ˈroði]	Mary ate the pomegranate.
K	[o ˈmanos ˈekopse to ˈroði]	Manos cut the pomegranate.
L	[i ˈmeri ˈeftise to ˈroði]	Mary spat the pomegranate.

[9] (fam) set was used in the familiarization phase and it was the same for all the participants.

Appendix E: Fillers

Filler ID	Picture Item	IPA transcription	Gloss	Target
FamF1	Eraser	[afto eðo ine ɣoma]	Is it (an) eraser?	Yes, it is.
FamF2	Spoon	[afto eðo ine sfugari]	Is it (a) sponge?	No, it is (a) spoon.
F1	pomegranate	[afto eðo ine roði]	Is it (a) pomegranate?	Yes, it is.
F2	apple	[afto eðo ine milo]	Is it (an) apple?	Yes, it is.
F3	light bulb	[afto eðo ine laba]	Is it (a) light bulb?	Yes, it is.
F4	yarn	[afto eðo ine laba]	Is it (a) yarn?	Yes, it is.
F5	spoon	[afto eðo ine kutali]	Is it (a) spoon?	Yes, it is.
F6	marker	[afto eðo ine markaðoros]	Is it (a) marker?	Yes, it is.
F7	(music) note	[afto eðo ine nota]	Is it (a) note?	Yes, it is.
F8	pencil	[afto eðo ine molivi]	Is it (a) pencil?	Yes, it is.
F9	lima	[afto eðo ine lima]	Is it (a) nail file?	Yes, it is.
F10	scissor	[afto eðo ine laba]	Is it (a) light bulb?	No, it is (a) scissor.
F11	knife	[afto eðo ine molivi]	Is it (a) pencil?	No, it is (a) knife.
F12	honey	[afto eðo ine roði]	Is it (a) pomegranate?	No, it is honey.
F13	oil	[afto eðo ine markaðoros]	Is it (a) marker?	No, it is oil.
F14	pomegranate	[afto eðo ine ɣoma]	Is it (an) eraser?	No, it is (a) pomegranate.
F15	apple	[afto eðo ine sfugari]	Is it (a) sponge?	No, it is (an) apple.
F16	light bulb	[afto eðo ine nima]	Is it (a) yarn?	No, it is (a) light bulb.
F17	Spoon	[afto eðo ine milo]	Is it (an) apple?	No, it is (a) spoon.
F18	marker	[afto eðo ine lima]	Is it (a) nail file?	No, it is (a) marker.

Appendix F: Praat script

```
# This script needs to be in the same folder as the wav files and the TextGrids

path$ = ""
sformat$ = ".wav"
gformat$ =".TextGrid"
wordsTier = 1
sylTier = 2
rowN = 0

#Create a table to save all the measures
tableName$ = "GISA_Praatdata"
#create table
Create Table with column names: "'tableName$'", 0, "Participant word syllable word_F0min word_F0max word_F0span word_duration syl_F0min syl_F0max syl_F0span syl_duration"


# Open all files in a directory

Create Strings as file list: "txtgrdObj", "'path$'*'gformat$'"
allFiles = Get number of strings
for gfile to allFiles
	selectObject: "Strings txtgrdObj"
	gridName$ = Get string: gfile
endfor
```

```
#open  each WAV file in the folder (directory)
for ifile to allFiles
	selectObject: "Strings txtgrdObj"
	filename$ = Get string: ifile
	namePrefix$ = filename$ - "'gformat$'"
	Read from file: "'path$''namePrefix$''sformat$'"
	soundname$ = selected$ ("Sound", 1)
	#open TextGrid
	selectObject: "Strings txtgrdObj"
	Read from file: "'path$''filename$'"
	tgridname$ = selected$ ("TextGrid")

	#get the values using the TextGrid
	interWords = Get number of intervals: wordsTier
	interSyl = Get number of intervals: sylTier

	#Getting durations: words
	for i to interWords
		wlabel$ = Get label of interval: wordsTier, i

		if  wlabel$ <> ""
			#get word duration
			wStart = Get starting point: wordsTier, i
			wEnd = Get end point: wordsTier, i
			wDur = wEnd - wStart

			#get F0 min; max; span: words
```

```
        selectObject: "Sound 'soundname$'"
        pObj$ = To Pitch: 0, 65, 500
        wf0max = Get maximum: wStart, wEnd, "Hertz", "Parabolic"
        wf0min = Get minimum: wStart, wEnd, "Hertz", "Parabolic"
        wf0span = wf0max - wf0min

        #add values to table
        selectObject: "Table 'tableName$'"
        Append row
        rowN = rowN + 1
        Set string value: rowN, "Participant", namePrefix$
        Set string value: rowN, "word", wlabel$
        #Set string value: rowN, "syl", slabel$
        Set numeric value: rowN, "word_F0min", wf0min
        Set numeric value: rowN, "word_F0max", wf0max
        Set numeric value: rowN, "word_F0span", wf0span
        Set numeric value: rowN, "word_duration", wDur

        selectObject: "TextGrid 'tgridname$'"
    endif
endfor
#Getting durations: syllable
for i to interSyl
    slabel$ = Get label of interval: sylTier, i

    if  slabel$ <> ""
        #get word duration
```

```
                sStart = Get starting point: sylTier, i
                sEnd = Get end point: sylTier, i
                sDur = sEnd - sStart

                #get F0 min; max; span: syllable
                selectObject: "Pitch 'namePrefix$'"
                sf0max = Get maximum: sStart, sEnd, "Hertz", "Parabolic"
                sf0min = Get minimum: sStart, sEnd, "Hertz", "Parabolic"
                sf0span = sf0max - sf0min

                #add values to table
                selectObject: "Table 'tableName$'"
                Append row
                rowN = rowN + 1
                Set string value: rowN, "Participant", namePrefix$
                Set string value: rowN, "syllable", slabel$
                Set numeric value: rowN, "syl_F0min", sf0min
                Set numeric value: rowN, "syl_F0max", sf0max
                Set numeric value: rowN, "syl_F0span", sf0span
                Set numeric value: rowN, "syl_duration", sDur

                selectObject: "TextGrid 'tgridname$'"

            endif
        endfor

endfor
```

Appendix G: R Formula & Performance and Comparisons of Models

Significance for all the models comparisons: *** p < 0.001; ** p < 0.01; * p < 0.05

F0max at word level

```
1. Empty model: word_F0max ~ 1 + (1 | Participant) + (1 | word)
## # Indices of model performance
##      AIC |      BIC | R2_conditional | R2_marginal |  ICC |  RMSE
## ----------------------------------------------------------------
## 18086.26 | 18108.39 |           0.81 |           0 | 0.81 | 28.80
```

```
2. Model with main fixed effects: word_F0max ~ 1 + Condition + Gender +
SubOrObj + Group + (1 | Participant) + (1 | word)
## # Indices of model performance
##      AIC |      BIC | R2_conditional | R2_marginal |  ICC |  RMSE
## ----------------------------------------------------------------
## 17967.87 | 18023.20 |           0.82 |        0.69 | 0.43 | 28.35
```

Comparison between model 1 and 2

```
##          npar   AIC   BIC  logLik deviance  Chisq Df Pr(>Chisq)
## Model 1     4 18086 18108 -9039.1    18078
## Model 2    10 17968 18023 -8973.9    17948 130.38  6  < 2.2e-16 ***
```

```
3. Final model: word_F0max ~ 1 + Condition + Gender + SubOrObj + Group +
Condition:SubOrObj + SubOrObj:Group + Gender:SubOrObj + Condition:Gender +
Condition:Group + (1 | Participant) + (1 | word)
## # Indices of model performance
##      AIC |      BIC | R2_conditional | R2_marginal |  ICC |  RMSE
## ----------------------------------------------------------------
## 17907.09 | 18023.27 |           0.83 |        0.70 | 0.44 | 27.72
```

Comparison between model 2 and 3

```
##          npar   AIC   BIC  logLik deviance  Chisq Df Pr(>Chisq)
## Model 2    10 17968 18023 -8973.9    17948
## Model 3    21 17907 18023 -8932.5    17865 82.785 11   4.26e-13 ***
```

Duration at word level

```
1. Empty model: word_duration ~ 1 + (1 | Participant) + (1 | word)
## # Indices of model performance
##      AIC |      BIC | R2_conditional | R2_marginal |  ICC | RMSE
## ---------------------------------------------------------------
## -3617.46 | -3595.33 |           0.35 |           0 | 0.35 | 0.09
```

```
2. Model with main fixed effects: word_duration ~ 1 + Condition + Gender +
SubOrObj + Group + (1 | Participant) + (1 | word)
## # Indices of model performance
##      AIC |      BIC | R2_conditional | R2_marginal |  ICC | RMSE
## ---------------------------------------------------------------
## -3708.36 | -3653.03 |           0.38 |        0.17 | 0.26 | 0.09
```

Comparison between model 1 and 2

```
##          npar     AIC     BIC logLik deviance Chisq Df Pr(>Chisq)
## Model 1     4 -3617.5 -3595.3 1812.7  -3625.5
## Model 2    10 -3708.4 -3653.0 1864.2  -3728.4 102.9  6  < 2.2e-16 ***
```

```
3. Final model: word_duration ~ 1 + Condition + Gender + Group + SubOrObj +
Condition:SubOrObj + Gender:SubOrObj + SubOrObj:Group + Condition:Group +
(1 | Participant) + (1 | word)
## # Indices of model performance
##      AIC |      BIC | R2_conditional | R2_marginal |  ICC | RMSE
## ---------------------------------------------------------------
## -3855.78 | -3750.66 |           0.44 |        0.22 | 0.28 | 0.08
```

Comparison between model 2 and 3

```
##          npar     AIC     BIC logLik deviance  Chisq Df Pr(>Chisq)
## Model 2    10 -3708.4 -3653.0 1864.2  -3728.4
## Model 3    19 -3855.8 -3750.7 1946.9  -3893.8 165.42  9  < 2.2e-16 ***
```

F0min at word level

```
1. Empty model: word_F0min ~ 1 + (1 | Participant) + (1 | word)
## # Indices of model performance
##       AIC |      BIC | R2_conditional | R2_marginal |  ICC |  RMSE
## ------------------------------------------------------------------
## 17334.78 | 17356.90 |           0.81 |           0 | 0.81 | 23.48
```

```
2. Model with main fixed effects: word_F0min ~ 1 + Condition + Gender +
SubOrObj + Group + (1 | Participant) + (1 | word)
## # Indices of model performance
##       AIC |      BIC | R2_conditional | R2_marginal |  RMSE
## -----------------------------------------------------------
## 17241.94 | 17297.27 |                |        0.80 | 23.48
```

Comparison between model 1 and 2

```
##          npar   AIC   BIC  logLik deviance  Chisq Df Pr(>Chisq)
## Model 1     4 17335 17357 -8663.4    17327
## Model 2    10 17242 17297 -8611.0    17222 104.83  6  < 2.2e-16 ***
```

```
3. Final model: word_F0min ~ 1 + Condition + Gender + SubOrObj + Group +
Condition:SubOrObj + Gender:SubOrObj + (1 | Participant) +  (1 | word)
## # Indices of model performance
##       AIC |      BIC | R2_conditional | R2_marginal |  ICC |  RMSE
## ------------------------------------------------------------------
## 17118.71 | 17190.63 |           0.82 |        0.74 | 0.32 | 22.67
```

Comparison between model 2 and 3

```
##          npar   AIC   BIC  logLik deviance  Chisq Df Pr(>Chisq)
## Model 2    10 17242 17297 -8611.0    17222
## Model 3    13 17119 17191 -8546.4    17093 129.24  3  < 2.2e-16 ***
```

F0span at word level

```
1. Empty model: word_F0span ~ 1 + (1 | Participant) + (1 | word)
## # Indices of model performance
##      AIC |      BIC | R2_conditional | R2_marginal |  ICC |  RMSE
## -----------------------------------------------------------------
## 18546.28 | 18568.41 |           0.32 |           0 | 0.32 | 33.18
```

```
2. Model with main fixed effects: word_F0span ~ 1 + Condition + Gender +
SubOrObj + Group + (1 | Participant) + (1 | word)
## # Indices of model performance
##      AIC |      BIC | R2_conditional | R2_marginal |  ICC |  RMSE
## -----------------------------------------------------------------
## 18502.99 | 18558.32 |           0.33 |        0.12 | 0.24 | 32.91
```

Comparison between model 1 and 2

```
##          npar   AIC   BIC  logLik deviance  Chisq Df Pr(>Chisq)
## Model 1     4 18546 18568 -9269.1    18538
## Model 2    10 18503 18558 -9241.5    18483 55.284  6  4.063e-10 ***
```

```
3. Final model: word_F0span ~ 1 + Condition + Gender + SubOrObj + Group +
Condition:SubOrObj + Gender:SubOrObj + Condition:Gender + (1 | Participant) +
(1 | word)
## # Indices of model performance
##      AIC |      BIC | R2_conditional | R2_marginal |  ICC |  RMSE
## -----------------------------------------------------------------
## 18404.58 | 18487.57 |           0.37 |        0.16 | 0.26 | 31.95
```

Comparison between model 2 and 3

```
##          npar   AIC   BIC  logLik deviance  Chisq Df Pr(>Chisq)
## Model 2    10 18503 18558 -9241.5    18483
## Model 3    15 18405 18488 -9187.3    18375 108.42  5  < 2.2e-16 ***
```

F0max at syllable level

```
1. Empty model: syl_F0max ~ 1 + (1 | Participant) + (1 | word)
## # Indices of model performance
##      AIC |      BIC | R2_conditional | R2_marginal |  ICC |  RMSE
## -----------------------------------------------------------------
## 16644.06 | 16666.19 |           0.91 |           0 | 0.91 | 19.44
```

```
2. Model with main fixed effects: syl_F0max ~ 1 + Condition + Gender +
SubOrObj + Group + (1 | Participant) + (1 | word)
## # Indices of model performance
##      AIC |      BIC | R2_conditional | R2_marginal |  ICC |  RMSE
## -----------------------------------------------------------------
## 16536.39 | 16591.72 |           0.91 |        0.76 | 0.63 | 19.18
```

Comparison between model 1 and 2

```
##          npar   AIC   BIC  logLik deviance  Chisq Df Pr(>Chisq)
## Model 1     4 16644 16666 -8318.0    16636
## Model 2    10 16536 16592 -8258.2    16516 119.67  6  < 2.2e-16 ***
```

```
3. Final model: syl_F0max ~ 1 + Condition + Gender + SubOrObj + Group +
Condition:SubOrObj + Condition:Group + Condition:Gender + (1 | Participant) +
(1 | word)
## # Indices of model performance
##      AIC |      BIC | R2_conditional | R2_marginal |  ICC |  RMSE
## -----------------------------------------------------------------
## 16385.55 | 16485.14 |           0.92 |        0.76 | 0.65 | 18.34
```

Comparison between model 2 and 3

```
##          npar   AIC   BIC  logLik deviance  Chisq Df Pr(>Chisq)
## Model 2    10 16536 16592 -8258.2    16516
## Model 3    18 16386 16485 -8174.8    16350 166.84  8  < 2.2e-16 ***
```

Duration at syllable level

```
1. Empty model: syl_duration ~ 1 + (1 | Participant) + (1 | word)
## # Indices of model performance
##       AIC |       BIC | R2_conditional | R2_marginal |  ICC | RMSE
## -----------------------------------------------------------------
## -7623.20 | -7601.07 |           0.39 |           0 | 0.39 | 0.03
```

```
2. Model with main fixed effects: syl_duration ~ 1 + Condition + Gender +
SubOrObj + Group + (1 | Participant) + (1 | word)
## # Indices of model performance
##       AIC |       BIC | R2_conditional | R2_marginal |  ICC | RMSE
## -----------------------------------------------------------------
## -7708.84 | -7653.51 |           0.42 |        0.15 | 0.32 | 0.03
```

Comparison between model 1 and 2

```
##           npar     AIC     BIC logLik deviance Chisq Df Pr(>Chisq)
## Model 1      4 -7623.2 -7601.1 3815.6  -7631.2
## Model 2     10 -7708.8 -7653.5 3864.4  -7728.8 97.64  6  < 2.2e-16 ***
```

```
3. Final model: syl_duration ~ 1 + Condition + Gender + Group + SubOrObj +
Condition:SubOrObj + (1 | Participant) + (1 | word)
## # Indices of model performance
##       AIC |       BIC | R2_conditional | R2_marginal |  ICC | RMSE
## -----------------------------------------------------------------
## -7735.13 | -7674.27 |           0.43 |        0.16 | 0.32 | 0.03
```

Comparison between model 2 and 3

```
##           npar     AIC     BIC logLik deviance Chisq Df Pr(>Chisq)
## Model 2     10 -7708.8 -7653.5 3864.4  -7728.8
## Model 3     11 -7735.1 -7674.3 3878.6  -7757.1 28.29  1  1.044e-07 ***
```

F0min at syllable level

```
1. Empty model: syl_F0min ~ 1 + (1 | Participant) + (1 | word)
## # Indices of model performance
##      AIC |      BIC | R2_conditional | R2_marginal |  ICC |  RMSE
## -----------------------------------------------------------------
## 16184.69 | 16206.82 |           0.91 |           0 | 0.91 | 17.19
```

```
2. Model with main fixed effects: syl_F0min ~ 1 + Condition + Gender +
SubOrObj + Group + (1 | Participant) + (1 | word)
## # Indices of model performance
##      AIC |      BIC | R2_conditional | R2_marginal |  RMSE
## ----------------------------------------------------------
## 16095.57 | 16150.90 |                |        0.90 | 17.12
```

Comparison between model 1 and 2

```
##          npar   AIC   BIC  logLik deviance  Chisq Df Pr(>Chisq)
## Model 1     4 16185 16207 -8088.3    16177
## Model 2    10 16096 16151 -8037.8    16076 101.12  6  < 2.2e-16 ***
```

```
3. Final model: syl_F0min ~ 1 + Condition + Gender + SubOrObj + Group +
Condition:SubOrObj + Condition:Group + SubOrObj:Group + Gender:SubOrObj +
(1 | Participant) +  (1 | word)
## # Indices of model performance
##      AIC |      BIC | R2_conditional | R2_marginal |  RMSE
## ----------------------------------------------------------
## 15920.08 | 16025.20 |                |        0.91 | 16.25
```

Comparison between model 2 and 3

```
##          npar   AIC   BIC  logLik deviance  Chisq Df Pr(>Chisq)
## Model 2    10 16096 16151 -8037.8    16076
## Model 3    19 15920 16025 -7941.0    15882 193.49  9  < 2.2e-16 ***
```

F0span at syllable level

```
1. Empty model: syl_F0span ~ 1 + (1 | Participant) + (1 | word)
## # Indices of model performance
##      AIC |      BIC | R2_conditional | R2_marginal |  ICC |  RMSE
## ---------------------------------------------------------------------
## 14685.95 | 14708.08 |           0.36 |           0 | 0.36 | 11.81
```

```
2. Model with main fixed effects: syl_F0span ~ 1 + Condition + Gender +
SubOrObj + Group + (1 | Participant) + (1 | word)
## # Indices of model performance
##      AIC |      BIC | R2_conditional | R2_marginal |  ICC |  RMSE
## ---------------------------------------------------------------------
## 14659.75 | 14709.55 |           0.37 |        0.08 | 0.31 | 11.73
```

Comparison between model 1 and 2

```
##         npar   AIC   BIC  logLik deviance Chisq Df Pr(>Chisq)
## Model 1     4 14686 14708 -7339.0    14678
## Model 2     9 14660 14710 -7320.9    14642  36.2  5  8.664e-07 ***
```

```
3. Final model: syl_F0span ~ 1 + Condition + Gender + SubOrObj + Group +
Gender:SubOrObj + SubOrObj:Group + Condition:SubOrObj + Condition:Gender + (1
| Participant) + (1 | word)
## # Indices of model performance
##      AIC |      BIC | R2_conditional | R2_marginal |  ICC |  RMSE
## ---------------------------------------------------------------------
## 14608.02 | 14702.08 |           0.39 |        0.12 | 0.31 | 11.52
```

Comparison between model 2 and 3

```
##         npar   AIC   BIC  logLik deviance  Chisq Df Pr(>Chisq)
## Model 1     9 14660 14710 -7320.9    14642
## Model 2    17 14608 14702 -7287.0    14574 67.731  8  1.388e-11 ***
```