MASTER IN COGNITIVE SCIENCE AND LANGUAGE

MASTER THESIS

September 2024

The Influence of Individuals' Musical Training and Pragmatic Skills on the Perception of f0 Manipulation in a Foreign Language

by *Larisa Gerenčir*

Under the supervision of:

Wendy Elvira-García

University of Barcelona

Abstract: In speech perception studies, particularly of intonation, several factors have previously been noticed to influence one's perceptive skills and, accordingly, their ability to perceive smaller changes in intonation. While sociodemographic factors like gender and age have often been studied in this context, other factors like musical training and individual differences, particularly individual pragmatic skills, are gradually becoming more interesting for those studying speech perception. Playing an instrument is argued to involve many highly complex cognitive functions and affect neuroplasticity and importantly, it has been studied in connection with auditory processing. Furthermore, the study of the influence of individual skills on auditory processing shows a positive relation between higher pragmatic skills and perception. While both of these factors are proven significant, the research tends to focus on one of them, rarely comparing the two skills or examining which one has a higher influence on the perception of intonation. The present study addresses this gap by examining Croatian native speakers' perception of f0 stimulation in Catalan, a language they are unfamiliar with. 91 participants took a discrimination test in which they listened to 20 pairs of stimuli and had to answer whether the stimuli in each pair were the same. The analysis was done in two steps, first, through a confusion matrix and analysis of answers for each pair of stimuli, and second, by building three linear regression models in order to compare the relationship between musical training and pragmatic skills and the correct answers.

The results demonstrate a good overall precision, but lower recall score and establish the fourth pair of stimuli as a threshold in which the majority of participants observe the difference between the stimuli. The second model, which included musical training, but no pragmatic skills as a variable, has proven to be the best-performing one, showing that in this study musical training played a more significant role than pragmatic skills in individuals' ability to perceive changes of f0 in an unfamiliar language.

Keywords: speech perception, intonation, musical training, pragmatic skills

Acknowledgments

First and foremost, I would like to express my deepest gratitude to my supervisor, Dr Wendy Elvira-García, for her patience and guidance during my whole master's degree, but especially throughout this research process. I am fortunate to have had an opportunity to learn from such a creative and dedicated professor and her passion for linguistics inspires me always to keep learning.

I am deeply grateful to Dr Joan Borràs-Comes who contributed to my thesis greatly, both with his suggestions and advice and by kindly sharing the stimuli used in this research.

I am also grateful to everyone I had a chance to learn from. I met some amazing professors who taught me, inspired me, and shared advice and suggestions. It was enriching to attend classes and participate in discussions with such experts during my master's and bachelor's studies. Moreover, I am thankful to my teachers from my hometown secondary and music school who helped me with suggestions, corrections, and participant recruitment.

I would like to extend my sincere thanks to everyone else who participated in my survey, shared it, or contributed to it with their comments or suggestions. This research would be impossible without your help and I truly appreciate your time and patience. Special thanks go to my parents who shared my survey with everyone they could think of and helped me reach this number of participants which seemed impossible at the beginning of the process.

In addition, I am very grateful to everyone who supported me by listening to my concerns, complaints, and worries and sharing their advice and love. My biggest thanks go to my family who supported me in every step of my education. I am also lucky to have so many wonderful friends who support me regardless of the distance between us. My family and friends' support is the main reason I can be here right now and I am truly indebted to each one of them.

List of Tables

List of Figures

Table of contents

1 Introduction

Music and speech are both argued to be innate to humans and researchers from various fields have observed and explained different aspects of the strong relationship between these two systems. One such commonly discussed theory is that musical training affects and possibly facilitates the processing, understanding, and learning of language. This hypothesis is supported by neuroscientific research, which has demonstrated that similar underlying brain areas are activated in music and speech perception (Schön et al., 2004; Marques et al., 2007).

In speech perception, the similarity emerges from both the speech prosody and musical melody relying on f0 (pitch contour), spectral characteristics, amplitude, and duration (Schön et al., 2004). In particular, it is argued that perceptually musical melody and speech prosody share the same underlying level known as contour, which contains information about up and down pitch changes, regardless of their exact size. Additionally, melodies have a second level called interval representation, which consists of an exact ratio of pitch between successive tones (Fujioka et al., 2004). While contour can be processed by all individuals, including children and musically untrained, the understanding of interval is considered specific for the musically trained. This can explain musically trained individuals' ability to more effectively recognise pitch violations and other changes in pitch contour and interval patterns (Schon et al., 2004; Fujioka et al., 2004).

Nonetheless, in their research of French listeners' perception of f0 stimulation in Portuguese sentences, Marques et al. (2007) demonstrated that all speakers are sensitive to some prosodic changes, including in an unfamiliar language, but musicians performed better in small differences. Furthermore, there is evidence that between four and eight hours of psychoacoustic training effectively lowers the threshold of non-musicians to that of musicians (Micheyl et al., 2006). Thus, it is important to look into other cognitive abilities to understand more thoroughly the processes involved in prosody perception.

Consequently, various individual differences are increasingly studied in this context, with many studies demonstrating their importance in someone's ability to understand smaller violations or changes in prosody (Kong and Edwards, 2011). One of the most studied individual differences is pragmatic skills, which include a person's ability to use language appropriately in social interaction, non-verbal communication, and attention switching, among others. In recent studies, pragmatic skills are most commonly analysed with the help of the Autism-spectrum Quotient (AQ), a self-report questionnaire developed to observe "autistic" traits in the neurotypical population (Baron-Cohen et al., 2001). The results of previous studies have shown that individuals with higher AQ scores, i.e. poorer pragmatic skills, are less sensitive to prosody overall (Bishop, 2012). Other experiments also exhibited the relationship between the AQ score and intonation processing in tone languages (Jiang et al., 2015), minimal phonetic pairs based on the voiced-voiceless distinction (Stewart and Ota, 2008), etc.

On the whole, it is argued that both musical training and individual differences will affect one's ability to perceive intonational changes. Nonetheless, so far, research has mostly focused on musical training or pragmatic skills individually, and not much study comparing them to see which one might be crucial and whether the influence of one can outweigh the other has been conducted.

This dissertation will contribute to the study of phonetic perception as it will address this gap by examining individuals' (native speakers of Croatian) pragmatic skills alongside their musical training and observing how they affect the perception of f0 manipulation in Catalan, a language that listeners are unfamiliar with. These two languages are interesting for such study because they are not that commonly studied in the field of phonetic perception; what is more, since they are from different language families and not usual L2 choice, it can be assumed that the speakers of one have not had a lot of contact with the other language and thus are not very familiar with its sound and prosody. This study will attempt to answer the following questions using a perception experiment:

1. What is the threshold for Croatian native speakers to perceive the f0 differences that are phonological in Catalan?
2. Do musical training or individual pragmatic skills have a higher influence on one's ability to perceive small changes in f0 frequency?

The structure of this dissertation will be as follows. Firstly, a literature review will explore previous research in the field of phonetic perception with a focus on musical training and pragmatic skills. This will offer the background information necessary to understand the hypothesis and results of the current study. Secondly, the current experiment will be explained in detail, including the information on all the analyses conducted. Following this, the results of the study will be presented, including the visualisations and observations made for both of the research questions. Moreover, the discussion that follows will relate these findings to previous research in the field, highlighting their importance and future implications as well as potential issues that could be addressed in the future. Lastly, the conclusion will summarise this study, and its results and offer some possible future lines of research that relate to it.

1.1 Review of the Literature

1.1.1 Music

Music and language are commonly described as uniquely human and it is impossible to imagine a human civilization that does not use them daily. Hence, it comes as no surprise a lot of research also attempted to explain the relationship between the two systems and in which way they affect each other. A strong connection has been found between melody in music and prosody in language, as they both rely on the same underlying acoustic parameters. (Schön et al., 2004, p.342). Thanks to these underlying parameters, it is largely considered that those individuals with higher musical skills due to having musical training develop a higher sensitivity to prosodic changes and violations. Out of all the acoustic parameters that music and language share, fundamental frequency (f0) is generally considered to be a primary cue for intonation perception (Shang et al., 2022). As discussed in Section 1, interval representation, which is typical for music, but does not occur in speech, seems to play an important role in the musicians' ability to perceive small prosodic violations and due to this is a somewhat more commonly investigated feature in the study of prosody.

Fujioka and colleagues (2004) investigated the relationship between long-term musical training and automatic melodic processing via neuroimaging (Magnetoencephalography) and behavioural tests. They found clear differences in brain activation between the musicians and non-musicians as well as their outperformance in the behavioural task. Their results point to the particular influence that musical training has in the early automatic stages of processing abstract levels of pitch contour and interval patterns. However, interestingly, it has less effect on the detection of simple pitch changes.

In another research, Schon et al. (2004) examined whether extensive musical training facilitates pitch contour processing in language as well as it does in music. They compared the professional musicians and non-musicians' electrophysiological and behavioural data and found that musicians outperformed non-musicians in both. They also found similar brain electrical potentials, hence their findings contribute to the findings that similar neural systems and cognitive computations are involved in f0 processing in language and music. Such findings have also been noticed by Deguchi et al. (2012) who studied whether the same cognitive and perceptive abilities used in pitch detection are different between musicians and non-musicians. According to their findings, musicians showed a significantly lower threshold and smaller within-group variability in their perception.

While much of the study focuses on native speakers' or L2 learners' perception of a language, less is known about prosody perception in an unfamiliar language. Nonetheless, Marques et al. (2007) compared French musicians' and non-musicians' perception of f0 stimulation of the last word in a sentence in Portuguese, which they did not speak. Their results proved that even in an unfamiliar language, we are sensitive to some prosodic changes, but musicians perform better with the smaller difference.

1.1.2 Pragmatic Skills

While studies often focus on groups and in many instances, the within-group similarities have been demonstrated, individual skills are also gradually getting more into focus in the speech perception study. Kong and Edwards (2011) found that in within-category sensitivity in voicing pairs, about 25% of the participants perceived the stimuli categorically, and individual differences influenced the perception. Within individual differences, one's pragmatic skills are particularly interesting to the researchers as they are tightly related to our language and communication and are argued to be a good predictor of one's speech perception ability.

In the language of autistic people, one of the commonly occurring features is a deficit in prosody, i.e. some individuals are unable to process or produce certain prosodic segments accordingly (Walenski et al., 2006). Due to this, autistic traits have been connected to a poorer perception of prosody and a lot of research started measuring one's pragmatic skills in these terms. The majority of people will have some traits that are also attributed to those on the spectrum, the so-called "autistic traits" and with a higher proportion of them, it is expected for one's pragmatic skills to be lower. With the development of the Autism-spectrum Quotient (Baron-Cohen et al., 2002), this became more standardised as this questionnaire, which measures the autistic traits in the typical population, soon became the researchers' preferred way of measuring them.

Although pragmatic skills are being gradually studied, the current findings and understanding of how they affect one's perception are inconsistent and prone to changing and expanding.

AQ communication score showed significance in the perception of narrow and wide focus statements in Bishop (2012). The native English listeners with high AQ were found not only less sensitive to prosody overall but also differed from those with lower scores in how they used it. They argue that a possible explanation for such behaviour is that they actively shift attention away from the prosodic prominence because it requires heavy usage of attention resources. Attention switching has also been demonstrated as having an important role in Stewart and Ota (2008) who investigated whether the AQ score will affect phonetic perception in the voiced and voiceless pairs between the real and imagined words. They presented their English-speaking participants with pairs of words, one of which was a real word beginning with a consonant, and the other one imagined with its voiced/voiceless counterpart instead. Their results showed that in individuals with higher AQ scores, their perception was less affected by the lexical knowledge, i.e. they perceived the word as it was pronounced rather than choosing the real word.

On the other hand, Jiang et al. (2015) tested whether speaking a tonal language would compensate for the speech deficit of individuals with Autistic Spectrum Disorder (ASD). As it is commonly accepted that the speakers of tonal languages demonstrate enhanced pitch processing ability compared to the speakers of non-tonal languages, they compared the individuals with ASD and the typical population's perception of pitch in music and speech. Their results showed that individuals with ASD exhibit normal or superior melodic contour processing and impaired intonation processing in Mandarin speakers with ASD.

The available research offers some valuable insight into the influence that one's pragmatic skills have on speech perception, although not much research has dealt with how the AQ score would affect the f0 manipulation in a language. What's more, the research so far has been conducted with native speakers of L2 learners of a language and it is yet to examine whether pragmatic skills also influence the perception of utterances in a foreign language. Nonetheless, taking the findings from the research mentioned in this section, we could assume that higher AQ might affect the perception negatively, i.e. since the individuals with higher AQ are less sensitive to prosody, they should notice the f0 change later than those with lower AQ scores.

2 Methodology

2.1 Stimuli

Stimuli used for this experiment were taken from a previous study on Catalan pitch intonation, undertaken by Borràs-Comes and colleagues (Borràs-Comes et al., 2010). It consisted of a continuum of 19 frequencies created by manipulating f0 of the Catalan word *petita* [pə.'ti.tə] (‘little’-fem) in Praat (Boersma & Weenink, 2024) as represented in Figure 1. The original sentence was pronounced by a native Catalan speaker with a rising-falling contour L+H* L%. Borràs-Comes et al. (2010) argue that in Catalan, this contour is used for expressing statements, contrastive foci, or echo questions and that the sentence type will depend on the pitch excursion range. Hence, the stimuli that the participants are presented with contain all three of the sentence types. While the first stimulus in the pair will always remain the same, the second stimulus changes in each pair and the participants have to recognise whether they are hearing the same or different stimuli.

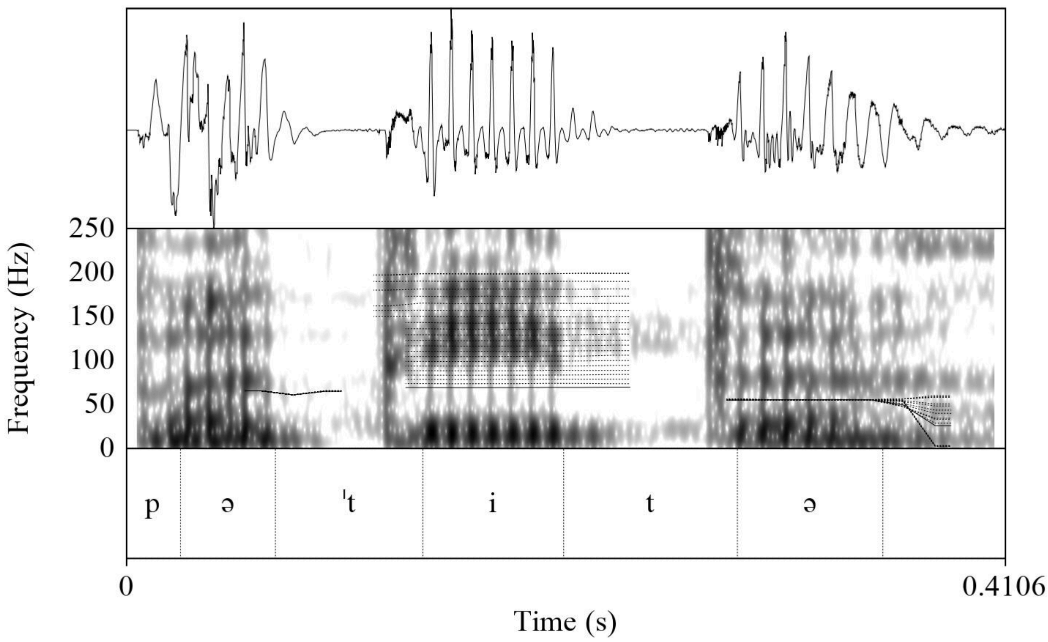

Figure 1. *Spectrogram, Waveform with Frequencies and Textgrid of the Stimuli*

2.2 Participants

Participants were recruited by word of mouth and by posting a survey link on social media. Participation was voluntary and none of the participants were financially or in any other way compensated for it. The criteria for participation were the following: (1) Native speakers of Croatian (2) Had to be at least fifteen years old, i.e. had finished primary school in the Croatian educational system (3) Had no prior knowledge of Catalan and did not understand it.

The participants were asked to answer some questions about their sociodemographic status, namely, gender, age, educational background and knowledge of foreign languages. A total of 93 participants were recruited for this experiment, however, two were excluded from the analysis because they spoke or understood the Catalan language. None of the participants reported any hearing or speech impairment that would prevent them from participating in this research.

The sociodemographic information of the participants is summarised in Table 1.

		N	**%**	**Mean**	**Range**	**SD**
Gender	Female	69	76%			
	Male	22	24%			
Age				28.39	45	10.76
Educational background	Primary School	3	3.3%			
	Secondary School	36	39.6%			
	Undergraduate	30	33%			
	Master	19	20.9%			
	PhD	3	3.3%			
Foreign language(s)	None (only Croatian)	11	12.1%			
	English	70	76.9%			
	German	1	1.1%			
	English and German	8	8.8%			
	English and Italian	1	1.1%			

Table 1. *Sociodemographic Information of Participants*

2.3 Procedure

The experiment was conducted via the online survey software Aidaform (Zakharenko, 2024). Before starting the survey, participants agreed to a consent form which contained information about the confidentiality, usage, and analysis of their data and answers. They were informed that the participation was voluntary, were given the author's contact details, and were instructed to contact them if they had any questions or wanted to withdraw their consent. Lastly, they were introduced to the structure, goals, and method of the survey and were instructed to fill it out in a quiet space, preferably using headphones.

The survey consisted of four sections. Firstly, participants answered questions about their sociodemographic background, which can be found in Table 1 above. Secondly, participants were asked about their musical training background. The participants who had any musical training or were self-taught were marked as Musicians, regardless of their musical proficiency. Besides that, they were asked what type of musical training they had and which group of instruments they played, all of which are summarised in Table 2.

Characteristic		**N**	**%**
Musicians	Yes	44	48.35%
	No	47	51.65%
Musical Training (Musicians Only)	Started Primary Music School	8	18.18%
	Finished Primary Music School	18	40.91%
	Finished Secondary Music School	7	15.91%
	Graduated from Music Academy	2	4.55%
	Self-taught	9	20.45%
Type of Instrument (Musicians Only)	Brass	2	4.55%
	Drums	1	2.27%
	Keyboard	12	27.27%
	Keyboard and Plucked string	5	11.36%
	Keyboard and Singing	2	4.55%
	Plucked string	10	22.73%
	Singing	1	2.27%
	Rhythm and Plucked string	1	2.27%

Woodwind	9	20.45%
Woodwind and Keyboard	1	2.27%

Table 2. *Information on Musical Knowledge of Participants*

Thirdly, they were asked to fill out the shortened version of the Autism Spectrum Quotient (Baron-Cohen et al., 2001) which consisted of 16 questions. As explained above, Attention Switching and Communication are found to be significant predictors of Perceptual Compensation (Yu, 2010), and to affect the perception of minimal phonetic pairs based on the voiced-voiceless distinction (Stewart and Ota, 2008). Thus, for this experiment, the questions were chosen from Attention to Detail (3), Attention Switching (3), and Communication (10) sections of the original questionnaire. The questions used in this survey were translated from original English to Croatian by the author and can be found in Appendix 1. They were answered using a Four Point Likert Scale, with possible answers being 1) Completely Agree, 2) Mostly Agree, 3) Mostly Disagree and 4) Completely Disagree.

Lastly, a perception test which consisted of 20 pairs of stimuli was conducted. In the first pair, the stimuli were the same, both being F1 and expressing statement contour with rising-falling pitch as it was pronounced by the native Catalan speaker. However, in the rest of the pairs, the first stimuli of the pair remained F1, but the second one was different, varying from F2 to F20. Before they started this part of the survey, participants were introduced to the nature of the questions and instructed to focus on the intonation of the two words to answer them accordingly. For each pair of audios, participants had to answer whether they thought the two audios were the same, where their options were: 1) Yes and 2) No.

2.4 Analysis

2.4.1 Features Analysed

Firstly, each participant's AQ score was computed in R (R Core Team, 2021) according to the scoring proposed by Baron-Cohen et al. (2001). "Completely Agree" and "Mostly Agree" were given one point in questions 1, 2, 5, 6, 8, 9, 14, and 16, whereas “Completely Disagree” and “Mostly Disagree” were given one point in questions 3, 4, 7, 10, 11, 12, 13 and 15. The data on AQ scores is summarised in Table 3.

	Mean	Range	SD
total_score	3.98	1 9	1.90

Table 3. *Mean, Range and SD values for the AQ score of the Participants*

Secondly, participants’ responses were assigned a value of 0 if they answered “No” and 1 if they answered “Yes”. The expected answers were marked in the same way, with the first pair given a value of 1, since it contains two stimuli of F1, and all the others given 0, as they differed. Participants’ responses were compared to the expected response to calculate the correct prediction for each question, where every True Positive (TP) and True Negative (TN) was assigned a value of 1, whereas False Positive (FP) and False Negative (FN) were assigned 0.

The third analysed feature was whether the participants were musicians or not. The features alongside their codes are summarised in Table 4.

Code	Feature	Scoring/Variables
Response	Participants' Responses to the Perception Task	1 for “Yes”; 0 for “No”
Expected	The correct response	1 for F1; 0 for F2-F20
PP	Correct prediction	1 for TP and TN; 0 for FP and FN
Question	A pair of stimuli	From F1 to F20
total_score	AQ score of a participant	From 0 to 16
GS	Whether the participant is musically trained	Yes or No

Table 4. *Summary of Analysed Features*

2.4.2 Statistical Analysis

All the data was analysed in R, version 4.0.4. (R Core Team, 2021). Firstly, to understand how successfully overall the participants identified the correct answer, a confusion matrix including precision, recall and F1 was calculated through the caret package (Kuhn, 2008). Precision gives us information on the proportion of correctly predicted TP out of all predicted positives and it is calculated through the following formula: *Precision = TP/(TP+FP)*. On the other hand, Recall (Sensitivity) contains information on the proportion of positive answers among those that were identified as positive and it is calculated as follows: *Recall = TP/(TP+FN)*. Additionally, as one of these functions can be significantly higher than the other, F1, an average of precision and recall, is calculated with the following formula: *F1 = (1+beta^2)*precision*recall/((beta^2 * precision)+recall)* (Kuhn, 2008).

Moreover, the distribution of answers for each individual question was calculated and visualised with ggplot2 (Wickham, 2016). This graph also helped to predict the threshold for recognising two stimuli as different by the majority of participants, which was later confirmed in the statistical analysis in the following step.

In the second step, the influence of different features on the correct identification was determined by employing three Generalised linear regression models (GLMMs; Table 5) built in the lme4 package (Bates et al., 2015). Given the data does not follow the Gaussian distribution, and that the correct answer was assigned a value of 1 and the incorrect one that of 0, a binomial family was employed in the calculation of the logistic regression. Model Combined included both the AQ scores of the participants and whether they were musicians, hence the results for Model Combined were calculated to look for any significant values with sjPlot (Lüdecke, 2024), including confirming the threshold which was predicted from the distribution of the answers. On the other hand, Model Music and Model Pragmatics focused on these two variables separately. Thus, to compare whether one influenced the results more, the three models were compared by plotting ROC curves with their predictions by pROC (Robin et al., 2011). The ROC curve is a graph that contains information on how well a model performs and as such it is a strong method to compare the influence of different features in this study. It is built from Sensitivity (look for Recall above) and Specificity, the proportion of TN out of all the identified negatives, calculated as *Specificity = TN / (TN + FP)* (Kuhn, 2008).

Model Combined	glm(PP~ Question + total_score + GS,data=test1, family = "binomial")
Model Music	glm(PP~ Question + GS,data=test1, family = "binomial")
Model Pragmatics	glm(PP~ Question + total_score ,data=test1, family = "binomial")

Table 5. *Generalised Linear Regression Models Built for the Test*

Lastly, the predicted values of the best-performing model were plotted for individual Questions to see how well the participants performed in each pair of stimuli and compare the musicians to non-musicians.

3 Results

The results of the Perception test showed that overall, achieving an accuracy of 0.833 with 95% CI between 0.815 and 0.8498. The participants were able to determine whether the stimuli were the same or different. As seen in Table 6, the participants did well on precision, which means they identified the true positive instances correctly. However, they performed poorer on recall, which suggests the existence of many false negatives, i.e. there were many instances where they misidentified different stimuli as the same. While this offers some insight into their performance, it does not help us understand which pairs of stimuli were difficult to differentiate or include any information on the influence of pragmatic skills or music on the results of this perception test.

Hence, to determine how the participants performed on each individual pair of stimuli, Figure 2 was plotted to visually represent the answer distribution. It shows the majority of the participants correctly identified the first pair of stimuli as the same. Moreover, they show somewhat unvaried responses from the fourth to the last question, where most of them correctly identified that the stimuli are different. The two pairs of stimuli that seem to have troubled them more were those in F2 and F3, which some identified correctly as different, but many misidentified as the same. That said, it can be concluded that F4 is a threshold in which most of the participants correctly identified the two stimuli as different.

	Precision	**Recall**	**F1**
Test 1	0.84615	0.20981	0.33624

Table 6. *Precision, Recall and F1 score for the Perception Test*

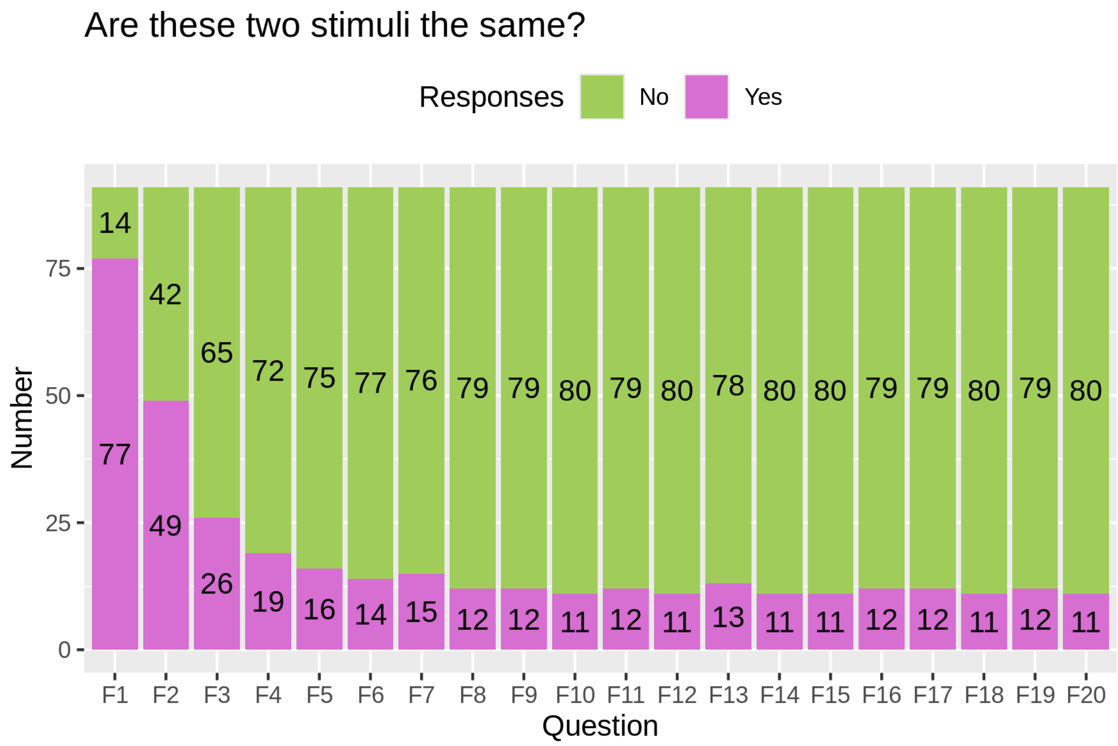

Figure 2. *Distribution of Answers for the Perception Task*

After building the models in glm, the analysis of Model Combined showed significant results in the intercept (F1), F2, and F3 and musically trained, as presented in Table 7. The significant p-values in F2 and F3 confirm the observations from the distributions of answers and demonstrate that these two questions received split perception and that the majority of the participants marked the threshold for identifying the stimuli as different. That said, we can confirm that F4, i.e. the third manipulated frequency, is a threshold in which the majority of the participants recognize the stimuli not to be identical.

Results for the First Perception Test

Predictors	*Odds Ratios*	*CI*	*p*
(Intercept)	**3.33**	**1.80 – 6.53**	**<0.001**
Question [F10]	1.34	0.56 – 3.23	0.512
Question [F11]	1.21	0.52 – 2.85	0.666
Question [F12]	1.34	0.56 – 3.23	0.512
Question [F13]	1.09	0.47 – 2.54	0.832

Question [F14]	1.34	0.56 – 3.23	0.512
Question [F15]	1.34	0.56 – 3.23	0.512
Question [F16]	1.21	0.52 – 2.85	0.666
Question [F17]	1.21	0.52 – 2.85	0.666
Question [F18]	1.34	0.56 – 3.23	0.512
Question [F19]	1.21	0.52 – 2.85	0.666
Question [F2]	**0.14**	**0.06 – 0.28**	**<0.001**
Question [F20]	1.34	0.56 – 3.23	0.512
Question [F3]	**0.44**	**0.20 – 0.91**	**0.029**
Question [F4]	0.68	0.31 – 1.47	0.326
Question [F5]	0.85	0.38 – 1.89	0.683
Question [F6]	1.00	0.44 – 2.29	1.000
Question [F7]	0.92	0.40 – 2.07	0.836
Question [F8]	1.21	0.52 – 2.85	0.666
Question [F9]	1.21	0.52 – 2.85	0.666
total score	1.01	0.95 – 1.08	0.712
GS [Yes]	**3.26**	**2.46 – 4.36**	**<0.001**
Observations	1820		
R^2 Tjur	0.103		

Table 7. *Results of the Model Combined*

The second part of the analysis intended to answer the starting question of this research, whether pragmatic skills or musical training influences the perception more. Furthermore, this model included both the pragmatic skills and musical training variables and the analysis indicates a significant p-value for the musically trained participants, while no significant value was marked for the AQ score, i.e. pragmatic skills. Such p-values in Model Combined suppose that while musical training influenced the participants' perception of frequency change, pragmatic skills did not play any significant role in it. To understand exactly what the influence of these factors is and how they compare to each other, two more models that focused on them individually were built; Model Music did not include the variable marking pragmatic skills and conversely, Model Pragmatics excluded the musical training. To compare them, the ROC curves for the three models were plotted together.

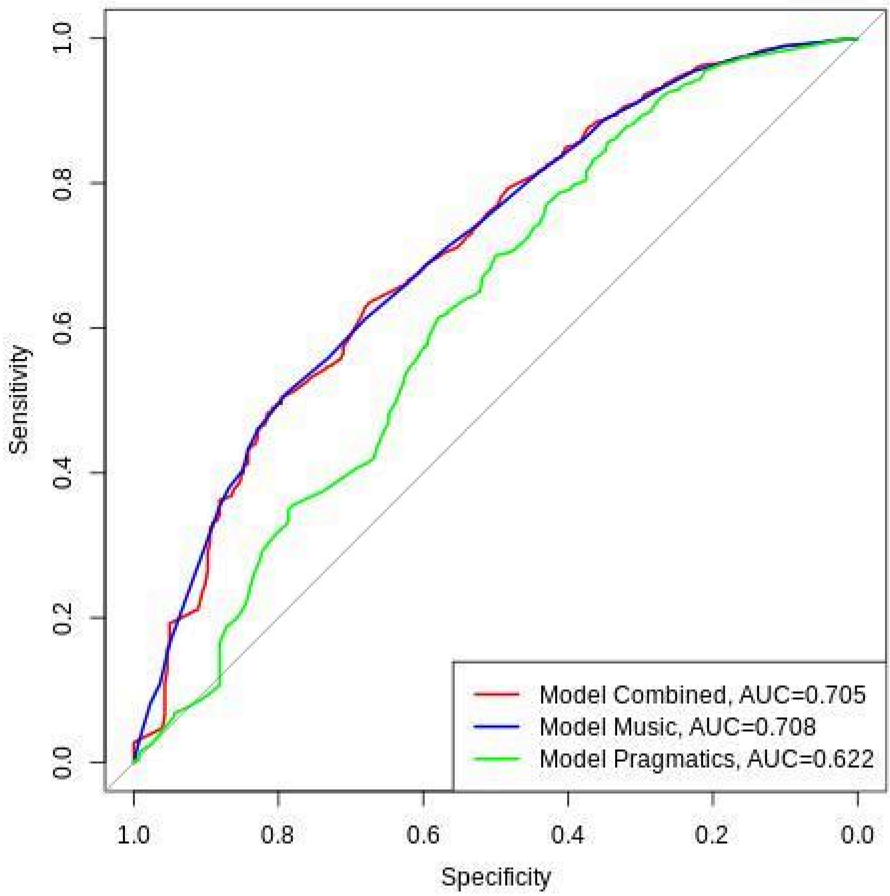

Figure 3. *ROC Curves for the Three Models*

Figure 3 contains ROC curves for the three Models and demonstrates not only the influence musical training has on perception, but also that its influence outweighs that of pragmatic skills. In this particular case, Model Music, which only accounts for the musical training, but not the pragmatic skills, is the best-performing one with the AUC=0.708, and due to this, it is taken in the following step of the analysis to compare the performance of musicians vs. non-musicians. Although both Model Combined and Model Pragmatics also surpass the random guessing line (AUC=0.5), Model Pragmatics which only contained the information on pragmatic skills, performed lower than Model Combined and Model Music, with AUC=0.622. Lastly, Model Combined with AUC=0.705

performed slightly poorer than Model Music which indicates that in the current study, accounting for the pragmatic skills of the participants not only does not improve the performance but in fact, lowers it slightly.

Finally, after observing the influence that musical training has on the perception of intonation change, a question of how it influenced each individual question in this study is imposed. For this, a graph that contained a prediction of Model Music as a y-axis and questions from F1 to F20 as an x-axis was plotted. As can be observed in Figure 4, the musicians outperformed non-musicians in each question of the Perception test, demonstrating an overall higher sensitivity to the frequency change. Moreover, it shows somewhat more even responses in musicians once they pass the threshold (F4, as explained above).

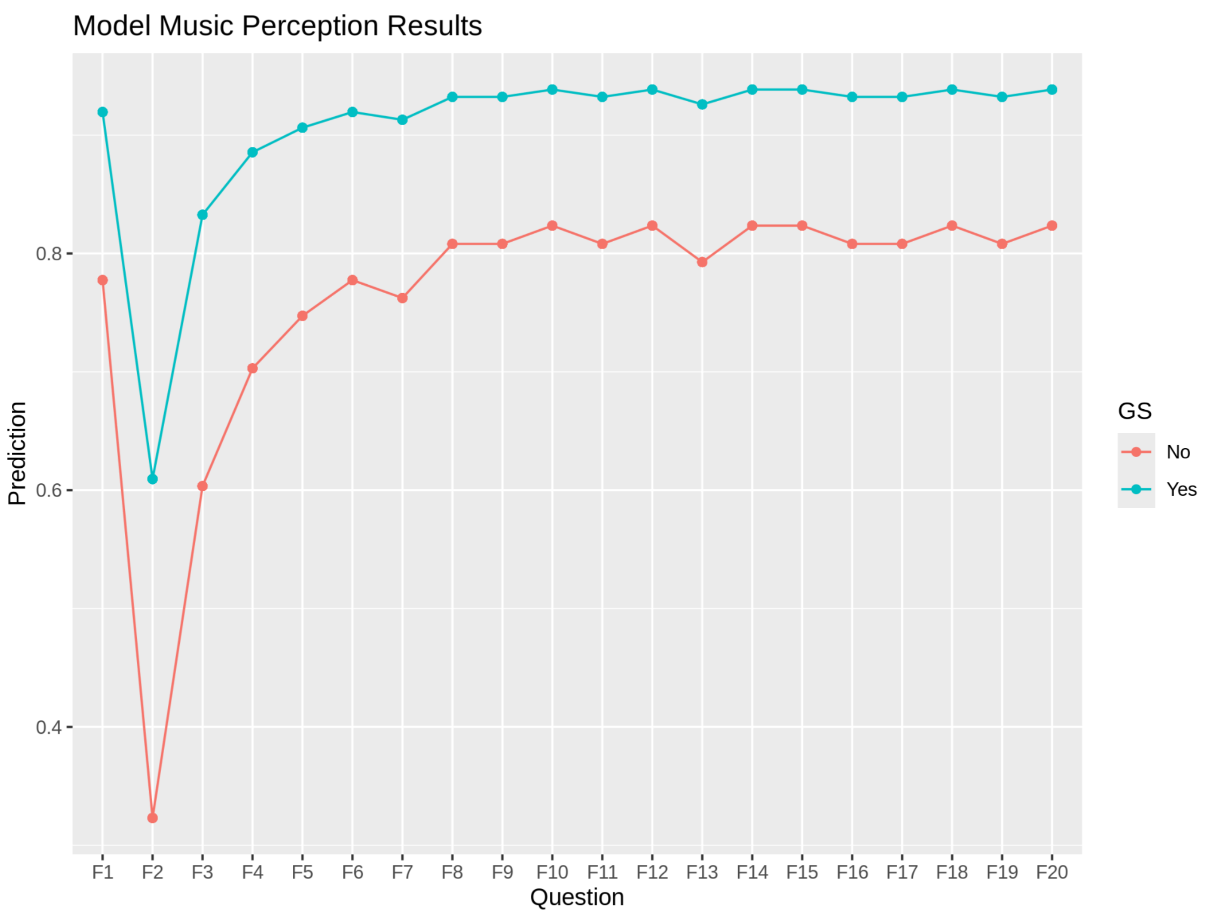

Figure 4. *The Comparison of Musicians' and Non-musicians' Performance in Model Music*

4 Discussion and limitations

4.1. Results discussion

The results obtained from this survey's analysis largely address the initial questions of this study and set a good ground for future comparisons between individual differences' influence on phonetic perception.

As determined in Section 3, the participants performed well overall in the Perception Task and demonstrated the ability to perceive a difference between the stimuli in a language that they had previously not been familiar with. As a rule, the participants struggled with the first two manipulations of the fundamental frequency and a significant p-value has been found for these two questions, demonstrating an unevenness in the answers. A general threshold has been established at F4, in which the majority of the participants (72) noticed the difference between the two stimuli.

Secondly, in the comparison between the influence that pragmatic skills and musical training have on one's ability to perceive manipulation of f0, the results contribute to the theory that musical training enhances perceptive skills. Contrary to Fujioka et al. (2004), in the current study, the musicians outperformed the non-musicians in each question of this perception task, showing that their training was not only valuable for perception of the early stages of pitch contour but also in determining simple pitch changes. Furthermore, similar to Deguchi et al. (2012) who found smaller within-group variability in musicians, the musicians in the current study also showed higher consistency in their results, especially once the threshold has been passed.

To answer the second research question, the glms built for this study helped determine that musical training has a higher influence on the participants' performance than their pragmatic skills did. Opposite to the starting hypothesis, lower pragmatic skills did not affect the perception negatively, and in fact, in this research, no significant difference was observed between the individuals based on their AQ score. Nonetheless, such results do not necessarily oppose the studies that have been previously done in the field (look at Section 1.1.2 above), since so far most of the research focused on native listeners or L2 learners of a language while in the current study, listeners were observing only one utterance which was isolated from the context and it was in an unfamiliar language. Thus, we can suppose that higher pragmatic skills have a higher influence on perception in one's native or L2 language and facilitate the prosody perception within the context, but not necessarily in an isolated situation.

Still, in the current study, including the AQ scores in the formula even slightly lowered the performance of the model, hence, the model which included only the musical training feature turned out to be the best-performing one.

4.2. Limitations and future lines of research

Although a significant correlation between the performance and the AQ score of the participants was not found in the current study, this experiment leaves some space for further exploration. As mentioned above, the original Autism-spectrum Quotient (Baron-Cohen et al., 2001) consists of fifty questions, only sixteen of which were used for this study. Even though the questions have been selected carefully, with earlier findings taken into account, the importance of pragmatic skills on intonation is only just being studied and more factors may influence it than previously thought. Consequently, either an extended AQ questionnaire or alternative ways of accounting for an individual's pragmatic skills might have yielded different results.

Additionally, having shown that musical training was a more important factor and the musicians achieved better in all questions of the perception task, the next step of the future analysis should be to compare the role that different levels and years of musical training have on the perception. Similarly, this dissertation does not explore the possible difference in the performance based on the type of instrument that the participants are playing. One research that included these two questions was conducted by Micheyl and colleagues (2006) and while they found no significance in the years of musical practice of their participants, they found that the musicians who played keyboard instruments, particularly piano, had a larger threshold in comparison to other musicians. Analysing these features might offer us even more insight into the phonetic perception of Catalan by Croatian speakers.

Furthermore, Borràs-Comes et al. (2010) in their research on Catalan speakers' perception of sentence type established the pitch range to be the main clue in differentiating between the statement and echo-question. Croatian speakers were also able to differentiate between the stimuli, but we do not know how well they would perform in the sentence type identification task. Such a study could offer valuable insight into the role f0 stimulation has cross-linguistically and could be applied in L2 teaching.

Aside from the two factors that this study examined, various other features have also been argued to affect one’s ability to correctly process and reproduce intonational patterns and other prosodical features in an L2. For instance, some of the most commonly mentioned ones are socio-demographic factors like gender, age, and educational background. While some of this information was also collected in the survey, they were not included in the analysis in this dissertation.

Moreover, numerous studies argue for the influence mother tongue experience has on learning prosody. In particular, a lot of attention has been paid to the study of tonal languages, as their speakers are generally considered more sensible to subtle changes in f0. Tonal languages, such as Mandarin Chinese, employ different f0 contours extensively to convey lexical meaning, alongside expressing post-lexical information, which is a primary function of f0 manipulation in non-tonal languages (e.g. Liu and Rodriguez, 2012; Shang et al., 2024). In Liu and Rodriguez’s (2012) article, Chinese speakers were found to have a steeper intonation identification than English speakers and for them, the intonation boundary was set closer to the level tone.

As the analysis showed, the threshold for the majority of the participants to perceive the difference between the stimuli was determined to be at F4, i.e. the third manipulated f0, which is arguably a good result and overall not a very high threshold. That said, Croatian is a pitch-accent language and its standard form alongside the majority of dialects makes use of a four-accent system (two falling and two rising) in which the f0 stimulation has an important role (Godjevac, 2006). Although establishing the threshold for the participants was one of the questions in this study, all of the participants spoke Croatian as their first language, and due to this, it is impossible to compare their performance to that of participants with a non-tonal mother tongue. Hence, a comparison between the non-tonal and tonal language would be necessary to understand this aspect more deeply.

5 Conclusion

This dissertation aimed to compare the influence musical training and pragmatic skills have on the perception of f0 manipulation in a language the listeners are unfamiliar with. As discussed in the introduction, previous research in the field proved both of these factors influence one's ability to perceive changes in prosody, and importantly for this study, small variations in intonation. However, previous research has been focused on one of the two features and there was no sufficient comparison between the two of them to determine which one influences the phonetic perception more.

For that reason, this study was conducted in the form of a survey that consisted of three crucial parts: information on musical training of the participants, Autism-spectrum Quotient, and Perception test. The study's goals were to determine the threshold in which the majority of Croatian native speakers would have recognised two stimuli as different and to compare the influence that their individual pragmatic skills and musical training have on their achievement. For that reason, an analysis of the data was done in two major steps. Firstly, an overall performance was observed and the distribution of the answers was accounted for visually. Secondly, three individual general regression models were built. The first one contained both AQ and musical training variables and it was analysed statistically to confirm the significant p-values. Subsequently, it was compared to the second model, which excluded pragmatic skills and the third model, which excluded the musical training variable in the ROC curves graph.

The threshold for perceiving two stimuli as different was found to be in F4, i.e. the third manipulated f0. Moreover, the results indicate a significant influence of musical training as musicians proved to be more sensitive to subtle frequency changes but also more consistent in their answers, achieving better results than non-musicians across every pair of stimuli. On the contrary, no significant relationship between the correct identification and AQ score was found, which indicates that in this study, individual pragmatic skills did not influence the participants' performance. Hence, the current study demonstrated musical training to be a more significant predictor of one's performance than their pragmatic skill.

These findings contribute to the research of phonetic perception in unfamiliar language, in particular, the influence that different extralinguistic factors have on the perception of intonation. Moreover, they set a good background for studying these questions in less-commonly represented Indo-European languages, in this case, Croatian and Catalan. Future research should build on these findings by expanding the number of factors included in the analysis and examining different patterns of intonation, or simply manipulating different aspects of prosody, such as rhythm, duration, speed, etc. Crucially, this line of research offers an important understanding of not only the phonetic perception, but the findings are also widely applicable in L1 and L2 teaching and speech therapy and rehabilitation.

6 References

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). The autism-spectrum quotient (AQ): Evidence from asperger syndrome/high-functioning autism, males and females, scientists and mathematicians. *Journal of autism and developmental disorders*, 31(1), 5-17

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

Borràs-Comes, J., Del Mar Vanrell, M., & Prieto, P. (2010). The role of pitch range in establishing intonational contrasts in Catalan. *Speech Prosody*. https://doi.org/10.21437/speechprosody.2010-151.

Bishop J. (2012) Focus, prosody, and individual differences in "autistic" traits: Evidence from cross-modal semantic priming. *UCLA Working Papers in Phonetics*, 111, 1–26.

Boersma, Paul & Weenink, David (2024). Praat: doing phonetics by computer [Computer program]. Version 6.4.13, retrieved 10 June 2024 from http://www.praat.org/

Deguchi, C., Boureux, M., Sarlo, M., Besson, M., Grassi, M., Schön, D., & Colombo, L. (2012). Sentence pitch change detection in the native and unfamiliar language in musicians and non-musicians: behavioral, electrophysiological and psychoacoustic study. *Brain research*, *1455*, 75–89. https://doi.org/10.1016/j.brainres.2012.03.034.

Fujioka, T., Trainor, L. J., Ross, B., Kakigi, R., & Pantev, C. (2004). Musical training enhances automatic encoding of melodic contour and interval structure. *Journal of Cognitive Neuroscience*, *16*(6), 1010–1021. https://doi.org/10.1162/0898929041502706

Godjevac, S. (2006). Transcribing Serbo-Croatian. *Prosodic typology: The phonology of intonation and phrasing*, 146-171.

Jiang, J., Liu, F., Wan, X., & Jiang, C. (2015). Perception of melodic contour and intonation in autism spectrum Disorder: evidence from Mandarin speakers. *Journal of Autism and Developmental Disorders*, *45*(7), 2067–2075. https://doi.org/10.1007/s10803-015-2370-4

Kuhn, M. (2008). Building Predictive Models in R Using the caret Package. *Journal of Statistical Software*, 28(5), 1–26. doi:10.18637/jss.v028.i05, https://www.jstatsoft.org/index.php/jss/article/view/v028i05.

Kong, E., & Edwards, J. (2011). Individual Differences in Speech Perception: Evidence from Visual Analogue Scaling and Eye-Tracking. *ICPhS*, 1126–1129. https://dblp.uni-trier.de/db/conf/icphs/icphs2011.html#KongE11

Liu, C., & Rodriguez, A. (2012). Categorical perception of intonation contrasts: Effects of listeners' language background. *The Journal of the Acoustical Society of America, 131*(6), EL427–EL433. https://doi.org/10.1121/1.4710836

Lüdecke, D. (2024). *sjPlot: Data Visualization for Statistics in Social Science*. R package version 2.8.16, https://CRAN.R-project.org/package=sjPlot.

Marques, C., Moreno, S., Castro, S. L., & Besson, M. (2007). Musicians detect pitch violation in a foreign language better than nonmusicians: Behavioral and electrophysiological evidence. *Journal of Cognitive Neuroscience*, *19*(9), 1453–1463. https://doi.org/10.1162/jocn.2007.19.9.1453

Micheyl, C., Delhommeau, K., Perrot, X., & Oxenham, A. J. (2006). Influence of musical and psychoacoustical training on pitch discrimination. *Hearing Research*, *219*(1–2), 36–47. https://doi.org/10.1016/j.heares.2006.05.004

R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J., & Müller, M. (2011). pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*, 12, 77.

Shang, P., Elvira-García, W. & Li, X. (2022). Cue weighting differences in perception of Spanish sentence type between native listeners of Chinese and Spanish. 644-648. 10.21437/SpeechProsody.2022-131.

Shang, P., Roseano, P., & Elvira-García, W. (2024). *Journal of Phonetics*, *102*, 101294. https://doi.org/10.1016/j.wocn.2023.101294

Schön, D., Magne, C., & Besson, M. (2004). The music of speech: Music training facilitates pitch processing in both music and language. *Psychophysiology*, *41*(3), 341–349. https://doi.org/10.1111/1469-8986.00172.x

Stewart, M. & Ota, M. (2008). Lexical effects on speech perception in individuals with "autistic" traits. *Cognition*, 109(1), 157-162. https://doi.org/10.1016/j.cognition.2008.07.010

Walenski, M., Tager-Flusberg, H., & Ullman, M.T. (2006). Language in autism. In Moldin, S.O., & Rubenstein, J.L.R., eds. *Understanding autism: From basic neuroscience to treatment* (pp. 175-203). Boca RatonFL: Taylor and Francis Books.

Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis. *Springer-Verlag New York*. ISBN 978-3-319-24277-4, https://ggplot2.tidyverse.org.

Zakharenko, A. (2024) AidaForm. Bonn, Germany. URL: https://aidaform.com

7 Appendix

Appendix 1.1. Questions Included in the AQ Questionnaire (English)

1. I often notice small sounds when others do not.
2. I tend to notice details that others do not.
3. I usually concentrate more on the whole picture, rather than the small details.
4. In a social group, I can easily keep track of several different people's conversations.
5. I find it easy to do more than one thing at once.
6. If there is an interruption, I can switch back to what I was doing very quickly.
7. Other people frequently tell me that what I've said is impolite, even though I think it is polite.
8. I enjoy social chit-chat.
9. When I talk, it isn't always easy for others to get a word in edgeways.
10. I frequently find that I don't know how to keep a conversation going.
11. I find it easy to "read between the lines" when someone is talking to me.
12. I know how to tell if someone listening to me is getting bored.
13. When I talk on the phone, I'm not sure I know when it's my turn to speak.
14. I am often the last to understand the point of a joke.
15. I am good at social chit-chat.
16. People often tell me that I keep going on and on about the same thing.

Appendix 1.2. Questions Included in the AQ Questionnaire (Croatian)

1. Često primjećujem male zvukove kada ih drugi ne primjećuju.
2. Sklon(a) sam primjećivati detalje koje drugi ne uočavaju.
3. Obično se više koncentriram na širu sliku, nego na male detalje.
4. U društvu lako mogu pratiti razgovore nekoliko različitih osoba.
5. Lako mi je raditi više od jedne stvari odjednom.
6. Ako dođe do prekida, mogu se vrlo brzo vratiti onomu što sam radila/o.
7. Drugi mi često govore da je ono što sam rekla/o nepristojno, iako ja smatram da je pristojno.
8. Uživam u (socijalnom) čavrljanju.

9. Kada govorim, drugima nije uvijek lako doći do riječi.
10. Često ne znam kako nastaviti razgovor.
11. Lako mi je "čitati između redaka" kada mi drugi nešto govore.
12. Mogu procijeniti kada sam dosadila/o mom sugovorniku.
13. Kada razgovaram na telefonu, nisam sigurna/siguran kada je moj red da govorim.
14. Često sam posljednja osoba koja shvati šalu.
15. Dobra/dobar sam u (socijalnom) čavrljanju.
16. Ljudi mi često govore da stalno pričam o jednoj te istoj stvari.