INVESTIGATING THE IMPACT OF SELECTIVE DIRECTIONAL AUDITORY ATTENTION ON EYE MOVEMENTS, PUPIL RESPONSES, AND AUDITORY PERCEPTION IN THE PRESENCE OF COMPETING SPEECH

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

SERAP YAĞMUR

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY
IN
THE DEPARTMENT OF INFORMATION SYSTEMS

SEPTEMBER 2023

INVESTIGATING THE IMPACT OF SELECTIVE DIRECTIONAL AUDITORY ATTENTION ON EYE MOVEMENTS, PUPIL RESPONSES, AND AUDITORY PERCEPTION IN THE PRESENCE OF COMPETING SPEECH

Submitted by **SERAP YAĞMUR** in partial fulfillment of the requirements for the degree **of Doctor of Philosophy in Information Systems Department, Middle East Technical** University by,

Date: **08.09.2023**

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

Name, Surname : Serap Yağmur

Signature :

ABSTRACT

INVESTIGATING THE IMPACT OF SELECTIVE DIRECTIONAL AUDITORY ATTENTION ON EYE MOVEMENTS, PUPIL RESPONSES, AND AUDITORY PERCEPTION IN THE PRESENCE OF COMPETING SPEECH

Yağmur, Serap

Ph.D., Department of Information Systems

Supervisor: Prof. Dr. Banu Günel Kılıç

Co-Supervisor: Prof. Dr. Kürşat Çağıltay

September 2023, 149 pages

This PhD thesis investigates how selective directional auditory attention affects eye movements and pupil responses, demonstrating that concentrating on stimuli from a single-ear enhances attentiveness and task performance. The study investigates how multi-talker speech interference, age, gender, and audio direction impact eye movements and pupil dilation. In a listening test, 38 subjects listened to short words with and without interfering speech. Interference noticeably affects pupils, prompting gaze towards sound sources and larger pupils during competing speakers, revealing a crucial aspect of the study. Moreover, this research uncovers an unexplored link between Perceptual Evaluation of Speech Quality (PESQ), Speech Intelligibility (SII), Signal-to-Noise Ratio (SNR) values for Turkish stimuli, and pupil dilation. Furthermore, this study examines whether presented stimuli involve familiar or unfamiliar words, probing emotional connotations effects on pupil dilation responses. The research establishes a link between speech quality and pupil dilation, indicating better performance with higher-quality auditory stimuli, regardless of lexical popularity or meaning. Ultimately, these findings underscore the utility of eye movement patterns and pupillometry as tools for capturing nuances in selective auditory attention. This holds profound implications for advancing hearing aid technology and improving communication abilities for individuals with autism spectrum disorder or hearing impairments.

Keywords: selective directional auditory attention, pupillometry, eye movement, competing speech, auditory perception

ÖZ

SEÇİCİ YÖNLÜ İŞİTSEL DİKKATİN GÖZ HAREKETLERİ, GÖZ BEBEĞİ TEPKİLERİ VE İŞİTSEL ALGI ÜZERİNDEKİ ETKİSİNİN REKABETÇİ KONUŞMA VARLIĞINDA İNCELENMESİ

Yağmur, Serap

Doktora, Bilişim Sistemleri Bölümü

Tez Yöneticisi: Prof. Dr. Banu Günel Kılıç

Ortak Tez Yöneticisi: Prof. Dr. Kürşat Çağıltay

Eylül 2023, 149 sayfa

Bu doktora tezi, seçici yönlü işitsel dikkatin göz hareketlerini ve gözbebeği tepkilerini nasıl etkilediğini araştırarak tek kulaktan gelen uyaranlara odaklanmanın dikkati ve görev performansını artırdığını göstermektedir. Çalışma, çoklu konuşmacılı ortamdaki konuşma girişimi, yaş, cinsiyet ve ses yönünün göz hareketleri ve gözbebeği genişlemesi üzerindeki etkisini incelemektedir. Bir dinleme testinde, 38 katılımcı, karışık konuşma olmadan ve karışık konuşma eşliğinde kısa kelimeleri dinledi. Müdahale özellikle göz bebeklerini etkiler, katılımcılar içgüdüsel olarak ses kaynaklarına bakış yönlendirir ve rakip konuşmacılar sırasında daha büyük göz bebekleri sergiler ve bu çalışmanın önemli bir yönünü ortaya çıkarmaktadır. Dahası, bu araştırma, Türkçe uyaranlar için Algısal Konuşma Kalitesi (PESQ), Konuşma Anlaşılabilirliği (SII), Sinyal-Gürültü Oranı (SNR) değerleri ile gözbebeği genişlemesi arasındaki ilişkiyi ortaya çıkarmaktadır. Ayrıca bu çalışma, sunulan uyarıların tanıdık veya tanıdık olmayan kelimeler içerip içermediğini ve duygusal çağrışımların gözbebeği genişlemesi tepkileri üzerindeki etkilerini araştırmaktadır. Araştırma, konuşma kalitesi ile gözbebeği genişlemesi arasında bir bağlantı kurarak, sözcüksel popülerlik veya anlamı ne olursa olsun daha yüksek kaliteli işitsel uyarıların daha iyi görev performansına yol açtığını göstermektedir. Sonuç olarak bu bulgular, göz hareketi desenleri ve pupillometriyi seçici işitsel dikkatin inceliklerini yakalamak için etkili araçlar olarak vurgulamaktadır. Bu, işitme yardımı teknolojisinin ilerlemesi ve otizm spektrum bozukluğu veya işitme engeli gibi durumlarla mücadele eden bireylerin iletişim becerilerinin geliştirilmesi konusunda derin etkilere sahiptir.

Anahtar Kelimeler: seçici yönlü işitsel dikkat, pupillometri, göz hareketi, rekabetçi konuşma, işitsel algı

Dedicated to my beloved family;

ACKNOWLEDGMENTS

First of all, I would like to extend my heartfelt gratitude to my supervisor, Prof. Dr. Banu Günel Kılıç, for her unwavering support, motivation, invaluable guidance, patience, and encouragement throughout the course of my research. Having her as my supervisor has been a tremendous pleasure and an honor. I am truly thankful for all the contributions she has made during my Ph.D. studies, which have greatly contributed to making this work productive and inspiring.

I am grateful to my co-advisor Prof. Dr. Kürşat Çağıltay, and to my thesis monitoring committee members, Assoc. Prof. Dr. Erhan Eren and Assoc. Prof. Dr. Mustafa Sert for their valuable feedbacks. I have learnt a lot from all of them. I would like to thank the examining committee members Prof. Dr. Altan Koçyiğit and Assoc. Prof. Dr. Betül Çiçek Çınar. I am also grateful Prof. Dr. Bilal Kırkıcı for allowing me to use FLE Eye Tracker Lab, thanks to him I completed the experiments.

I am grateful to Attorney Başak Akgün Akil and Attorney Kürşat Uçar for their contributions to record speech materials.

I am also deeply appreciative of the study participants; their involvement was indispensable for the successful completion of this research.

I acknowledge the support of The Scientific and Technological Research Council of Turkey (TÜBİTAK) BİDEB for 2211-E graduate student fellowship during my Ph.D. education.

I would like to thank all of my colleagues and friends, Ahmet Bodur, Melih Şırlancı, Gökhan Gönül, Efecan Yılmaz, Ayhan Serkan Şık.

TABLE OF CONTENTS

ABSTRACT iv
ÖZ v
ACKNOWLEDGMENTS vii
TABLE OF CONTENTS viii
LIST OF TABLES xii
LIST OF FIGURES xiii
LIST OF ABBREVIATIONS xv
CHAPTERS
CHAPTER 1 1
INTRODUCTION 1
1.1. The Purpose of the Study 2
1.1. Research Questions 2
1.2. Contributions of the Study 3
1.3. Organizations of the Thesis 4
CHAPTER 2 5
LITERATURE REVIEW 5
2.1. Speech Location 5
2.2. Speaker Gender 8
2.3. Environmental Characteristics 9
2.3.1. Noise and SNR (Signal to Noise Ratio) 9
2.4. Speech Intelligibility 11
2.4.1. Speech Intelligibility Measurement 12
2.4.2. Technical Measures of Speech Intelligibility 13
2.5. Speech Perception 19
2.6. Measurement Methods for Listening Attention 20

2.6.1. Eye Tracking and Pupillometry 21
2.6.2. Dual Task Paradigm 26
CHAPTER 3 29
METHODS AND EXPERIMENT DESIGN 29
3.1. Speech Material 29
3.2. Recording of Speech Material 30
3.3. Stimuli 31
3.4. Participants 36
3.5. Apparatus 36
3.5.1. EyeLink 1000 Tower 37
3.5.2. Software: Experiment Builder 38
3.6. Experiment Design 39
3.7. Task and Procedure 41
3.8. Data Analysis 43
3.9. Pilot Study 44
3.10. Data Preprocessing and Analysis 44
3.11. Ethics 45
3.12. Assumptions of the Study 45
CHAPTER 4 47
EFFECTS OF SELECTIVE DIRECTIONAL AUDITORY ATTENTION ON EYE MOVEMENTS AND PUPIL RESPONSES 47
4.1. Research Questions of the Experiment 47
4.2. Data Preprocessing and Analysis 47
4.3. Results 48
4.3.1. Eye Movement Related Results 49
4.3.2. Pupil Related Results 52
4.3.3. Overall Accuracy Results 59
4.4. Discussion 61
4.4.1. Audio Direction Effects 61
4.4.2. Participant Gender Effects 62
4.4.3. Interference Effects 63
4.4.4. Age Effects 63
4.5. Chapter Conclusion 64

CHAPTER 5 67
UNRAVELING AUDITORY ATTENTION AND PUPIL RESPONSES: INVESTIGATING DIFFERENCES BETWEEN HEARING AND LISTENING, SEMANTIC IMPACT, AND STIMULUS FAMILIARITY 67
5.1. Research Questions of the Experiment 67
5.2. Methods 67
5.2.1. Word Frequencies in Turkish 67
5.2.2. Emotional Sentiment Analysis 68
5.2.3. Word Embedding (Word2Vec) Analysis 69
5.3. Procedure 70
5.3.1. Clustering Data into 3 groups 70
5.4. Data Analysis 71
5.4.1. Pre-Data Analysis 71
5.5. Results 72
5.5.1. Pupil Dilation Correlation and Linear Regression Analysis Results 72
5.5.2. Peak Dilation Correlation and Linear Regression Analysis Results 74
5.5.3. Peak Latency Correlation and Linear Regression Analysis Results 76
5.5.4. Accuracy Correlation and Linear Regression Analysis Results 77
5.5.5. Pupil Dilation Linear Mixed Effect Model Results 79
5.5.6. Peak Dilation Linear Mixed Effect Model Results 85
5.5.7. Peak Latency Linear Mixed Effect Model Results 88
5.5.8. Overall Accuracy Linear Mixed Effect Model Results 91
5.6. Discussion 94
5.6.1. Effects of Speech Intelligibility Metrics 94
5.6.2. Word Frequency and Semantic Content Results 96
5.6.3. Hearing or Listening Effects 97
5.7. Session Conclusion 97
CHAPTER 6 99
GENERAL CONCLUSIONS 99
6.1. Practical Implications 102
3.1.1 103
6.2. Limitations and Future Work 103
REFERENCES 105

APPENDICES 123
7.1. APPENDIX A 123
7.2. APPENDIX B 146
CIRRICULUM VITAE 147

LIST OF TABLES

Table 1: Stimuli Presentation 32
Table 2: Presented stimulus which contains 10 Turkish three-character monosyllabic word 32
Table 3: Vector group of Speech Material 69
Table 4: Linear Mixed Effect Model of Pupil Dilation 79
Table 5: Interactions of Pupil Dilation Linear Mixed Effect Models 83
Table 6: Linear Mixed Effect Model of Peak Dilation 85
Table 7: Linear Mixed Effect Model of Peak Latency 88
Table 8: Linear Mixed Effect Models of Overall Accuracy 91
Table A.1: Derived Values of SII, SNR, PESQ 123
Table A.2: Frequency of use of each word 127
Table A.3: Sentiment analysis of stimuli 135

LIST OF FIGURES

Figure 1: Searching three-character monosyllabic Turkish words in kelimeler.net..... 30
Figure 2: METU SPARG (Spatial Audio Research Group) Lab at the Modeling and Simulation Center 31
Figure 3: Test environment..... 37
Figure 4: EyeLink 1000 Tower (Figure is taken from EyeLink 1000 User Manual) 38
Figure 5: Experiment Builder 39
Figure 6: View of the instruction screen shown to participants..... 39
Figure 7: View of the attention clue given to the participants 40
Figure 8: Displayed marker before playing the audio..... 40
Figure 9: Blank screen displayed while the audio is played 41
Figure 10: View of the screen where participants type the words they recall 41
Figure 11: A picture from Experiment..... 42
Figure 12: (a) Effect of Audio Direction on horizontal eye positions (Fixation X) for left, bilateral and right audio directions. (b) Effect of Audio Direction, Interference and Participant Gender on Fixation X. 50
Figure 13: Saccade X (horizontal rapid eye movement) results recorded at the onset of each stimulus according to the direction of the target speech and the presence of interference. ... 52
Figure 14: Baseline pupil sizes, calculated as the mean pupil size during the 1 s interval before the speech onset, according to Age Group and Participant Gender, showing that the Baseline Pupil Size decreases with age and female participants have larger baseline pupil sizes than male participants. 52
Figure 15: Pupil Size results calculated as the mean pupil sizes during the 2 s presentation of the word listened to by each participant. (a) According to Audio Direction and presence of Interference for different Age Groups, showing that there was a significant difference in Pupil Size results according to Audio Direction and presence of Interference, and that female participants had larger Pupil Sizes than male participants, and that as the age of the male participants increased, their pupil size decreased. (b) According to presence of Interference for Participant Gender, showing that the presence of Interference causes a significant increase in the Pupil Size results (c) According to whether the participants Focused on the Competing Speaker or not and the Speaker Gender for Participant Gender, showing that the Pupil Size results of the participants who focused on the competing speaker instead of the target speaker differed according to the gender of the participant. 54
Figure 16: Pupil Dilation results, calculated by subtracting the baseline pupil size from the pupil size for each word and condition for all participants individually. (a) According to Audio Direction and presence of Interference for different Age Groups, showing that there was a significant effect of Audio Direction on the Pupil Dilation, Pupil Dilation results of male participants in the Age Group 19 to 25 were significantly greater than for the male participants in the Age Group 26 to 44, and the effect of Age Group on Pupil Dilation for female participants was not significant. (b) According to presence of Interference for Participant Gender, showing that in cases of Interference, both male and female participants' Pupil Dilation increased, and that the male participants' Pupil Dilation was larger than female

participants'. (c) According to whether the participants Focused on the Competing Speaker or not and the Speaker Gender for Participant Gender showing that there is a significant difference between the Pupil Dilation results of female participants who Focused on Competing Speaker instead of Focusing on Target Speaker regardless of the speaker gender. 56
Figure 17: Peak Pupil Size, defined as the participants' largest pupil size during the entire time each trial was being listened to, according to the Audio Direction and Interference for Participant Gender, showing the Peak Pupil Size was greater in the presence of Interference, and Peak Pupil Size results of female participants were significantly larger than the male participants, and the participants in the Age Group 19 to 25 were found to have Larger Peak Pupil Size results than the participants in the Age Group 26 to 44. 58
Figure 18: (a) Peak Pupil Size in mm according to the Peak Latency and Participant Gender showing that the female participants have larger Peak Pupil Size values and the Peak Pupil Sizes reached while listening to the 7th and 8th words were the largest (b) Count of the Peak Pupil Size cases according to the Peak Latency and Participant Gender, showing that the Peak Pupil Size was reached mostly while listening to the last words of each trial. 58
Figure 19: Overall Accuracy of the participants taken as the number of correct words, i.e., spoken by the target speaker (a) based on the Audio Direction, presence of Interference and Participant Gender, showing that in the absence of interference, participants were more successful in listening situations where the sound was given from the left than from the right. (b) based on the presence of Interference, Speaker Gender and Participant Gender, showing that the male participants were more successful than female participants, and this difference was more when the speaker was male. 60
Figure 20: Frequency of the order of the words in the stimuli that were written, showing that the first 3 and the last 3 words of the stimuli were entered more than the other words of the stimuli by the participants. 60
Figure 21: Response type effect on pupil dilation 80
Figure 22: Target word vector effect on pupil dilation 81
Figure 23: Target word PESQ effect on pupil dilation 82
Figure 24: Competing word PESQ effect on Pupil Dilation 82
Figure 25: Target word vector effect on peak dilation 86
Figure 26: Target word PESQ effect on peak dilation 87
Figure 27: Target word SII effect on peak dilation 87
Figure 28: Target Word PESQ effect on Peak Latency 90
Figure 29: Target word SII effect on peak latency 90
Figure 30: Target word vector effect on overall accuracy 93
Figure 31: Target word SII effect on overall accuracy 94

LIST OF ABBREVIATIONS

AI: Articulation Index
dB: Decibel
dBA: A-weighted Decibel
DRT: Diagnostic Rhyme Test
ERP: Event-Related Potentials
HLS: Hearing Loss Simulations
Hz: Hertz
KHz: Kilohertz
LLR: Log Likelihood Ratio
LPP: Late Positive Potential
MOS: Mean Opinion Scores
MRT: Modified Rhyme Test
NA: Not Applicable
NH: Normal Hearing
PESQ: Perceptual Evaluation of Speech Quality
RT: Response Time
SII: Speech Intelligibility Index
STI: Speech Transmission Index
SNR: Signal Noise Ratio
SNRseg: Segmental Signal-to-Noise Ratio
VR: Virtual Reality
WSS: Weighted Spectral Slope

CHAPTER 1

INTRODUCTION

The process of communication involves not only perceptual factors such as the ability to hear, but also cognitive factors. Hearing constitutes a sense, whereas listening entails a skill that demands focused attention and deliberate intention to fully grasp the conveyed information. Understanding speech in noisy environments becomes more difficult and requires effort. When listening attention decreases, social interaction declines and this can lead to a decrease in quality of life and increase in stress.

Factors that affect listening attention continue to be explored by researchers. In our study, we conducted experiments related to the "Cocktail Party" topic, which is frequently mentioned in the literature. As many researchers have stated, communicating in dynamic environments like a cocktail party requires more listening effort due to attention processes. In our daily lives, we are used to hearing various sounds from different directions, and despite the noise in our environment, we still manage to communicate. So, how do we communicate with others and distinguish each sound when exposed to so much noise? Researchers have been trying to answer this question for many years by focusing on the sound propagation environment, the direction of sounds, the differences between listeners, and the attention given to sounds as well as their importance, through various experiments. Many studies focus on selective listening, i.e., listening to only the designated target speaker in the presence of interfering speech. Some of these studies investigate the effects of sound, the acoustic characteristics of the environment, and the source position, while others measure and interpret the participants' reactions (body, head, and eye movements). In some studies, these two factors are considered together.

In this study, we aimed to measure the participants' attention to sound direction by collecting data on their eye movements and pupil dilation while they listened to single and multi-talker sounds. Additionally, we investigated the relationship between speech intelligibility and listening attention by examining how the stimuli presented to the participants affected their pupil dilation. For this purpose, we calculated the SII and PESQ scores of the stimuli and tested different levels of speech content. We also analyzed the relationship between the SNR levels of the stimuli and pupil dilation. Furthermore, we explored the impact of the presented stimuli being known or unknown words on the participants.

The participants were presented with single-syllable three-letter Turkish words as stimuli. These stimuli were produced by a female and male native Turkish speaker. The stimuli sometimes consisted of content produced by a single speaker, and other times were composed of different word groups produced simultaneously by a female and male participant in order to observe the effects of interference. Participants were given visual information prior to each trial indicating which speaker they should focus on. The relationship between focusing and pupil dilation was investigated.

1.1. The Purpose of the Study

The purpose of the study is to examine the effects of participant gender, age, competing speech, and sound location on listening attention. The study aimed to measure auditory attention to sound location using eye movements and pupil dilation, taking into account gender differences. The stimuli, consisting of words spoken by either a male or female speaker, were presented to the right ear only, the left ear only, or both ears. The hypothesis was that participants would tend to look to the right when the target sound came from the right and to the left when it came from the left, thus horizontal and vertical eye gaze fixations and saccades were examined. Pupil dilation was recorded to measure attention to sound location, and baseline pupil size, peak pupil size, and pupil latency were calculated. The study also tested whether speech intelligibility decreased in the presence of a competing talker in the multi-talker environment, leading to an increase in processing load, larger pupil dilation, and decreased performance of the participants. To the best of our knowledge, so far no study examined the relationship between selective directional listening attention and eye movement and pupil dilation. This study, aims to fill this gap.

The study also aimed to investigate the effect of signal-to-noise ratio (SNR), speech intelligibility based on the Speech Intelligibility Index (SII) and the Perceptual Evaluation of Speech Quality (PESQ) on listening attention. Measuring listening attention could be useful in understanding differences in the impact of hearing loss across different environments and in quantifying attentional demands for tasks, such as driving. The study also aimed to investigate whether the meaning and semantic content of auditory stimuli influence a person's listening attention, and if so, how this influence is reflected in changes in the size of their pupils during listening tasks. Additionally, measuring and modeling listening attention could be useful for optimizing the balance between speech intelligibility and listening comfort in hearing aids.

1.1. Research Questions

This Ph.D. thesis aims to address critical research questions concerning auditory attention, perceptual evaluation metrics, eye movements, pupil dilation, and contextual factors. The following research questions guide the investigation:

Chapter 4 focuses on addressing the following research questions:

1. Does the level of auditory attention have a relation with pupil size in processing speech?
2. Does interference have an effect on auditory attention, and can this effect be determined through changes in pupil size?
3. Does listener age and gender affect auditory attention?
4. Does the source location of the sound affect listening to a task?

Chapter 5 delves into exploring the following research questions:

5. Can we investigate the difference between hearing and listening using eye movements?
6. What is the effect of SII (Speech Intelligibility Index), PESQ (Perceptual Evaluation of Speech Quality), and SNR (Signal-to-Noise Ratio) values on auditory attention and pupil responses?
7. Does the semantic content have an impact on auditory attention, and how does this impact manifest in pupil dilation?
8. How does the familiarity of stimuli presented within the study influence auditory attention and pupil dilation?

In Chapter 4, we will comprehensively examine and provide insights into auditory attention dynamics and its relationship with various factors such as interference, spatial location, and listener gender, utilizing physiological responses like changes in pupil size as indicators. In Chapter 5, we will shift the focus towards exploring the influence of semantic content, perceptual evaluation metrics, and familiarity of stimuli on auditory attention, leveraging eye movements and pupil dilation for a deeper understanding of perceptual and cognitive aspects of auditory perception.

1.2. Contributions of the Study

This Ph.D. thesis contributes significantly to the field of auditory attention by exploring a multidimensional approach that integrates perceptual evaluation metrics, eye movements, pupil dilation, and cognitive aspects. The research findings and methodologies presented herein advance our understanding of auditory attention mechanisms and their intricate relationship with various influencing factors. The primary contributions of this research can be summarized as follows:

- Unveiling the Interplay of SII, PESQ, and SNR values on auditory attention and pupil responses.
- Explanation of selective auditory attention through eye movements and pupil dilation.
- Demonstration of changes in pupil dilation in cases of interference.
- Exploring the influence of listener age and gender on auditory attention through eye movements and pupil dilation.
- Examining semantic content and word familiarity impact on auditory attention through eye Movements and pupil dilation.

1.3. Organizations of the Thesis

The thesis has six chapters. Chapter 1 provides information about the study and presents the study's objectives.

Chapter 2 presents a literature review, which includes the methods of listening attention, eye-tracker, and pupillometry used in listening studies, as well as speech intelligibility, SII, SNR, and PESQ and speech perception.

Chapter 3 provides a detailed description of the user experiment, including the research method, experiment design, devices and instruments used, a pilot study, participant information, and data collection procedure.

Chapter 4 examines the impact of selective directional auditory attention on eye movements and pupil responses.

Chapter 5 presents the study of the effect of signal-to-noise ratio (SNR), speech intelligibility calculated using the SII (Speech Intelligibility Index), PESQ (Perceptual Evaluation of Speech Quality), and semantic content on listening attention.

Finally, Chapter 6 concludes the thesis by summarizing its contributions and practical implications, and outlining future work.

CHAPTER 2

LITERATURE REVIEW

2.1. Speech Location

There are usually several competing acoustic sources present in an environment during any communicative situation (Kidd et al., 2005). Competing sounds can be classified as nonspeech noises and other talkers' speech (Reel & Hicks, 2012). In these situations, selective auditory attention plays a significant role in ignoring competing sounds and focusing on the target speech (Koch et al., 2011; Reel & Hicks, 2012; Yost et al., 1996) Most studies on the cocktail party effect consider "selective attention" tasks, in which the listener is to attend to one target source in the presence of competing sources (Yost et al., 1996). To illustrate, while listening to your friend's interesting story at a cocktail party, you suppress or ignore the competing acoustic sources, such as the background noise and other conversations (Koch et al., 2011).

It was previously shown that the differences among the spatial locations of sound sources affect the selective attention in the presence of competing speech (Freyman et al., 2004; Reel & Hicks, 2012). If the direction of the target speech is known, communication becomes relatively easier than otherwise (Kitterick et al., 2010). The studies of Kidd et al. (2005) and Brungart & Simpson (2007) showed that while listening to two different talkers, prior knowledge of the target speech's location improves the listening performance. Also, as the degree of the interference of the competing speech signal decreases, the performance of auditory attention increases (Reel & Hicks, 2012). However, despite the studies conducted on the cocktail party phenomena, it is not very well understood how to determine which direction the listener is paying attention to.

Numerous dichotic listening studies have been conducted to examine the selective directional auditory attention (Cherry, 1953; see Pashler, 1998, for review). In these studies, participants focus on sounds played in one ear in dichotic listening and ignore the other ear's sounds. Moreover, auditory attention switching has been investigated by Koch et al. (2011), where a novel selective listening model was designed to investigate the mechanisms of voluntary switching in selective auditory attention employing an auditory word categorization task. According to this model, in each trial,

two different single number words were played simultaneously. Each number word was spoken to each ear by distinct speakers, one female and one male. At the beginning of the trial, each gender was represented with a color, which was provided as the cue of the target speaker during the listening. After the trial, subjects had to indicate whether the number spoken by the target speaker was smaller or greater than five. As a result, it was found that auditory attention switches are costly, i.e., require extra task completion time (Koch et al., 2011).

Vertegaal et al. (2001) found that people tend to look at the talker they are listening to in multi-talker conversations. According to this study, the reason for this is to get additional information by looking at the speaker's mouth movements, gestures, gaze direction, and facial expression to increase speech intelligibility. As Yost et al. (1996) stated, target localization plays a significant role in focusing on the target speech. Thus, head movement and eye movement measurements have been used to investigate sound localization. Kim et al. (2013) examined the role of people's movements in a localization task. They found that people turn their heads towards the sound source. Vertegaal et al. (2001) stated that users' gaze direction could be a cue for predicting who was talking to whom. Another study conducted by Hendrikse et al. (2018) found that people have movement strategies to localize a target. According to their study, people look at the target speaker. If the target speaker is not in sight, people are most likely to turn their heads toward the target speaker. Tessendorf et al. (2011) conducted an experiment in which body and eye movements were recorded to show that these movements could improve hearing. They used electrooculography (E.O.G.) to record eye movements and found that eye movement analysis performed better than body localization analysis in terms of accuracy.

One earlier study of eye movements in selective listening tasks was conducted by Gopher (1973). He measured eye movements by the Electro-Oculograph (E.O.G.) method and recorded only the horizontal eye movements. He stated that eye movement could be a sign to observe people's behavior in a dichotic listening test of focused attention. A large eye movement frequently occurs in the direction of selective monitoring of the ear (Gopher, 1973). One of the participants who knew this phenomenon said that " ... If I look to the right, I can listen to the right ear. If I look to the left, I can listen to the left ear ....".

We used the eye tracker methodology to investigate the attention role in sound localization. Eye-tracking is a method to determine the point of a person's gaze or track eyes' motion concerning their head's position. Previous studies have attempted to capture the eye and facial expression features to predict attention in different conditions (Bakry et al., 2020; Ghosh et al., 2015; Pavlov & Kotchoubey, 2017). Eye gaze, head pose, and facial action details are the key markers for predicting interaction (Braga et al., 2016; Kwok, 2018). Kawahara et al. (2018) stated that eye gaze was the most important property to anticipate the interaction's identification. Also, eye gaze was more efficient than nodding, laughing, and backchannels to expect the mental states (Putze et al., 2018). Braga et al. (2016) examined participants' task-based eye movements during the auditory task with the absence of visual stimuli by recording vertical and horizontal gaze to find an association between effective listening and eye-movement changes. The results showed that it was possible to detect auditory attention using eye movement (Bakry et al., 2020; Braga et al., 2016).

Another study using an eye tracker system related to attention was conducted by Salvaggio et.al (2019). They investigated the attention shifts while processing numbers mentally. They asked the participants to compare the given numbers (20 to 70) with the reference number (45) when looking at a blank screen or a screen with colorful pictures. They stated that participants' gaze shifted leftward and downward while listening to smaller numbers than the reference. Moreover, small numbers were linked to the left side of space, while large numbers were linked to the right side.

On the other hand, Hendrikse et al. (2018) investigated the influence of different visual cues on the head and eye movement in a multi-talker listening environment. They stated that the eye movement was more when compared to the head movement. According to them, the reason was that people shifted their eyes to the speaker to predict the target speaker. Some researchers used saccadic movements instead of eye gaze patterns. The saccade is described as rapid movements of each eye from one gaze location to another (Lee et al., 2002; Leigh & Zee, 1991). Lee et al. (2002) stated that saccadic eye movements and head rotation in the same direction co-occur.

In the auditory attention studies, apart from the eye movement measurements, the use of pupillometry became popular after the innovative publications by Beatty, Kahneman, and colleagues (Kahneman & Beatty, 1966; Laeng et al., 2012). Pupil size has been presented to be responsive to cognitive resource distribution changes in many different tasks that show auditory stimuli (Kahneman & Beatty, 1966). Zekveld et al. (2018) reported that pupil size was sensitive to task demand and emotional responses, tiredness, and attention. While many studies were trying to measure internal attention due to the presence of sound, many others tried to measure the intentional or voluntary attention by presenting dichotic auditory stimuli (for review, see Zekveld et al., (2018)). In these studies, that are also called task-evoked, pupil response was dealt with as a change occurring in the pupil diameter phase. The researchers interpreted the increase in pupil size while the participant's task and engagement goes on (Granholm et al., 1996). Baseline pupil size was associated with arousal or task engagement (Aston-Jones & Cohen, 2005). Some of the studies that were done using pupil dilation are reviewed below.

In Jin et al. (2018), pupil dilation, eye blink, and eyelid opening have been used as point of reference to measure the attention level. Briefly, as the complexity of speech increases, the pupil size increases as well. Beatty, (1982) states that the pupil response is correlated with attention. One study related to the auditory attention used pupil dilation to detect auditory attention switches (McCloy et al., 2017). They measured pupil dilation, target detection sensitivity, and participants' reaction time while listening to a concurrent speech stream. They reported that attention switching is related to pupil response. They stated that switching attention between talkers increased pupil dilation. Koelewijn et al. (2015) conducted three experiments to show that the pupil response is affected by the prior knowledge of the target speech location, target onset time, and who will speak. Their results showed that location uncertainty increased the pupil dilation response. They concluded that communicating in dynamic environments like a cocktail party, requires more listening effort because of the attention processes. In other words, normal hearing participants' pupil dilation responses are smaller when knowing the location of the target speech than when its location is uncertain (Koelewijn et al., 2015). Besides, Koelewijn et al. (2017) state

that the attention to the speaker and the speaker's location enhances the speech intelligibility. According to Kidd et al. (2005) and Kitterick et al. (2010) knowing where to listen improves speech perception performance.

Nowadays, because of the pandemic, people tend to interact online instead of physical interaction (Siyaev & Jo, 2021). Therefore, virtual reality (VR) has become more popular with metaverse applications. VR has been used in many different fields, such as education, medicine, games, disaster management, neurodegenerative disorders assessment, and cognitive psychology studies (Arias et al., 2021; Imaoka et al., 2020; Shimizu et al., 2021). Eye tracking technology has been integrated into the Virtual Reality (VR) technology with the help of head-mounted display (HMD). This improvement makes it possible to measure real-time eye gaze, saccade, fixation, and pupil diameter during watching VR animations (Imaoka et al., 2020). For example, Imaoka et al. (2020) measured saccadic eye movements with a simple environment to evaluate whether the HMD-based VR technology can be used as a suitable tool. Enders et al. (2021) investigated gaze behaviour during navigation and visual search in a naturalistic virtual environment. They measured number of fixations, dwell time (duration of fixation), saccade rate and blink rate. They stated that even in complex virtual environments, the eye tracker data reliably showed the task-related gaze behaviour and was strongly related to the target condition. Zheng et al. (2020) used an eye tracker integrated into the VR environment and the pupil diameter to classify emotions. Similarly, Chen et al. (2017) conducted an exploratory study to measure pupil dilation response in different types of VR experience. They investigated how emotional VR content causes pupil dilation and whether pupil dilation was a reliable measure of emotional response. They claimed that pupil dilation was affected by the VR environments. Eckert et al. (2021) measured the pupil dilation to evaluate the cognitive load during a VR experience. They expressed that the eye tracker was a suitable device to measure the cognitive load in the VR environment.

The majority of these studies focused mostly on visual search, visual attention, and cognitive load in VR environments and recorded user responses while measuring eye movements with eye-tracking devices. As people's interaction in virtual environments is increasing, studies on selective auditory attention become more important to create a baseline which can be compared with that in a VR environment. For example, in Enders et al. (2021) study, the participants were expected to listen to numbers during a math task and continue navigation in a virtual environment at the same time, but it was seen that some participants paused the navigation and were only paying attention to the listening.

2.2. Speaker Gender

Considering the literature, some factors have been studied more when compared to others. Although listener-related factors, such as the age, gender, and hearing status were emphasized to be important, the impact of these individual differences was not investigated enough by the researchers (Peelle, 2018; Pichora-Fuller et al., 2016). Voiers et al. (1973) stated that they only tested male speakers. Koelewijn et al. (2015) conducted a study on pupil size response and listening effort using one female and one male speaker, but they did not mention the effect of the speaker's gender in the results.

McGarrigle et al. (2017) used speech stimuli recorded by a native English-speaking female. Only five studies (Burley et al., 2017; Dabbs, 1997; Gilzenrat et al., 2010; Gingras et al., 2015; and Partala & Surakka, 2003) in the literature deal with the gender differences using male and female pupil size data. As Zekveld et al. (2018) reported, there must be more research for these remaining questions, and also, the impact of the listener factor should be investigated. Moreover, Zekveld et al. (2018) found that most studies in the literature have investigated the relationship between pupil dilation and listening effort, and some have included memory as well. However, to the best of our knowledge, so far no study examined the relationship between selective directional listening attention and eye movement and pupil dilation. This study, aims to fill this gap.

2.3. Environmental Characteristics

In the literature, there are studies related to the relationship between environmental characteristics and speech understanding. In the study conducted by Hällgren et. al. (2005), an examination was carried out to assess the impact of hearing aids on perceived cognitive effort, speech perception, and tasks involving demanding speech comprehension across two distinct settings: quiet and noisy environments. The findings revealed that participants reported reduced cognitive effort when employing hearing aids in comparison to instances where hearing aids were not utilized, particularly in situations characterized by background noise.

2.3.1. Noise and SNR (Signal to Noise Ratio)

In verbal communication, background noise such as a passing car or a noisy air conditioner can negatively impact the quality of speech signal and lead to decreased speech intelligibility and increased listener fatigue. In our daily lives, we are faced with many sounds that make our communication more difficult. Listening in a noisy environment requires much effort, and as a result, the understanding of the speech reduces. The capacity to comprehend speech is intricately linked with cognitive attributes like working memory, processing speed, and phonological processing (Gatehouse et al., 2004). Research has indicated that individuals with typical hearing, spanning both pediatric and adult populations, tend to exhibit a higher propensity to discontinue challenging tasks when confronted with auditory stimuli in noisy settings, as opposed to tranquil environments (e.g., Cohen et al., 1980).

The impact of noise on speech understanding is measured by signal-to-noise ratio (SNR). The study conducted by Larsby (2005) has demonstrated the detrimental impact of noise on the precision and speed of speech processing. The central focus of this study typically encompasses auditory tasks, primarily involving activities such as word recognition conducted under conditions of tranquility or amidst noise at a predefined signal-to-noise (SNR) ratio, which serves as an evaluative parameter for assessing listening exertion and effort. SNR, known to be associated with speech comprehension (Plomp, 1986), holds relevance in influencing the efficacy of hearing aid interventions, as observed in earlier studies (Walden et al., 2005; Wu and Bentler, 2010).

The signal-to-noise ratio (SNR) is a commonly used measure for evaluating the quality of speech, but it has some drawbacks. It is simple to calculate, but requires both distorted and clean speech samples. The classic definition of SNR may not accurately reflect the perceived quality of speech, because it averages the relationship over the entire signal, even though the energy of speech fluctuates. To address this issue, the SNR can be calculated in segments and then averaged, which is called segmental SNR. This method takes into account the impact of noisy segments on overall quality and is more closely aligned with how humans perceive speech quality. Segments with high energy and no noticeable noise do not dominate the overall perception of quality, but the presence of noisy segments stands out and lowers the overall quality.

SNR is calculated as the ratio of the signal energy to the noise energy, expressed in decibels:

$$SNR = 10\log_{10}\frac{\sum_{i=1}^{N} x^2(i)}{\sum_{i=1}^{N}(x(i)-y(i))^2} \quad \textbf{(1)}$$

The SNR is calculated by comparing the original and processed speech samples, represented by *x(i)* and *y(i)* respectively, over a series of samples indexed by *i*, with *N* being the total number of samples. Instead of using the whole signal, the Segmental Signal-to-Noise Ratio (SNRseg) calculates the average SNR values of shorter segments, typically 15 to 20 ms in length. It is given by:

$$SNR_{seg} = \frac{10}{M}\sum_{m=0}^{M-1}\log 10 \sum_{i=Nm}^{Nm+N-1}\left(\frac{\sum_{i=1}^{N} x^2(i)}{\sum_{i=1}^{N}(x(i)-y(i))^2}\right) \quad \textbf{(2)}$$

The SNRseg uses the segment length, represented by *N*, and the number of segments, represented by *M*, to calculate the average SNR values. SNRseg provides more accurate results for waveform encoders, but it gives poor results when used with vocoders.

The first study to examine SNRs in real-world speech-listening situations was conducted by Pearsons et al. (1977), who found that as the noise level increased, the speech level increased and the SNR decreased. Pearsons et al. (1977), have shown that when noise level falls below 45 dBA, speech level in the listener's ear remains at a constant 55 dBA. As the noise level increases, speech level increases linearly by 0.6 dB. When noise reaches 70 dBA, SNR decreases to 0 dB with 15.5% of the measurements having SNRs below 0 dB. SNR has a significant impact on speech understanding (Plomp 1986) and hearing aid outcomes (Walden et al. 2005; Wu and Bentler 2010).

The data from the study conducted by Pearsons et al. (1977) has been widely used to determine the SNR in speech-related tests for individuals with normal hearing and with hearing loss. However, the participants in the study were adults with normal hearing. In a more recent study, Smeds et al. (2015) estimated the SNRs in real-world environments encountered by hearing aid users with moderate hearing loss. The results were not completely in line with those reported by Pearsons et al. The SNRs were 3 to 5 dB higher than those reported by Pearsons et al., especially in situations with low-level noise. Smeds et al. (2015) stated that the difference in results could be due to the difference in research participants and the methods used for collecting and analyzing the recordings.

2.3.1.1. SNR Level

In the literature, different levels of Signal-to-Noise Ratio (SNR) are used for different purposes and research studies. A higher SNR value indicates a better quality of the speech signal and less impact of the background noise. The choice of the SNR level used in a study depends on various factors, including the research question, the population being studied, and the specific goals of the study. Some studies may use higher SNR levels to assess speech understanding in ideal conditions, while others may use lower levels to simulate real-world environments where background noise is present. In any case, it is important to understand the SNR level used in a study to accurately interpret the results and their implications.

2.4. Speech Intelligibility

The main criterion for evaluating the effectiveness of a speech system is its speech intelligibility, or the degree to which speech can be understood by a listener (Letowski, 2017). Speech intelligibility is a measure of the amount of a message that has been extracted from the recognized phonemes, or the smallest units of speech (Côté, 2011). It indicates the extent to which words and sentences can be understood (Viswanathan and Viswanathan 2005). Speech intelligibility is typically expressed as the percentage of speech units that are received and understood correctly.

The purpose of the study conducted by Rennies (2014) was to investigate the relationship between listening effort, speech intelligibility, and predictions of the speech transmission index (STI) in noisy and reverberant environments. The results showed that as STI increased, listening effort always decreased, indicating that while intelligibility and listening effort are related, listening effort may be a better measure of speech perception in situations where intelligibility is already high. The goal of the study was to improve methods for predicting speech perception in rooms, which could be useful in practice (Rennies, 2014).

2.4.1. Speech Intelligibility Measurement

2.4.1.1. Speech Intelligibility Rating

Rating is a method of evaluating objects or stimuli based on their perceived quality or value using a verbal or numeric scale of successive categories. The speech material used in a rating assessment can be phrases, sentences, or connected speech that is representative of the specific application (Letowski, 2017).

The ITU P.800 standard, proposed by the International Telecommunications Union (ITU) in 1996, includes a 5-point Absolute Category Rating (ACR) scale that is commonly used to measure speech quality. The results from multiple subjects are typically combined to compute the Mean Opinion Score (MOS) for a given sample of speech (Sharma et. al, 2016).

The MOS (Mean Opinion Score) test is the most commonly used statistical rating system for evaluating speech intelligibility. In a MOS test, participants rate their perception of a speech sample on a 5-point scale from 1 (bad) to 5 (excellent) with 5 indicating “excellent” quality and 1 indicating “unsatisfactory” or “bad” quality. The ratings are averaged to obtain a final score (ITU-T P.800, 1996).

2.4.1.2. Perceptual Test

The intelligibility of speech is important for comprehending the amount of speech information that is both available and audible to a receiver during disruptive conditions that could prevent comprehension of a spoken message. Subjective (perceptual) and objective (technical) tests are used to measure speech intelligibility.

Measuring perceptual speech intelligibility can be challenging, as it requires subjective judgments from listeners. One common method is to use speech recognition software to analyze the speech signal and determine how well it was understood by the listener. Another method is to use a questionnaire or survey to ask listeners directly about their understanding of the speech. Perceptual tests of speech intelligibility directly measure the percentage of correctly understood speech items. These tests are used in international and national standards, and involve presenting human listeners with various forms of speech (e.g., phonemes, syllables, words, phrases, sentences, or paragraphs) and collecting their recognition responses. The speech material used in these tests can be live, recorded, or synthetic. In addition to assessing overall speech intelligibility, the collected data can be analyzed to provide information about the types of errors made and specific deficiencies encountered. However, the validity of these tests is limited to the specific speech material and complexity level used (Letowski,2017)

There are many different speech intelligibility tests that are available in various languages and formats, including nonsense syllables, single words, and sentences. Of these, sentences are the most realistic representation of communication. Nonsense

syllables and single words can be used to assess confusion and analyze the transmission of information. In general, sentence tests have shown steeper slopes (between 0.15 and 0.25 per dB) in intelligibility functions compared to nonsense syllables or single-word tests. (Brand, 2008)

There are several factors that can affect perceptual speech intelligibility, including the listener's age, hearing ability, and familiarity with the speaker's accent or dialect. In addition, the quality of the speech signal itself can affect perceptual speech intelligibility, such as the level, clarity, and intelligibility of the speech.

Speech intelligibility typically refers to how well speech can be understood. There are several objective measures used to evaluate intelligibility, including the speech intelligibility index (SII) and the perceptual evaluation of speech quality (PESQ). (Nathwani, 2017). The SII model calculates the average amount of speech information available to a listener, with a value of 0 representing complete unintelligibility and a value of 1 representing perfect intelligibility.

According Loizou (2007), in general, a method with higher SII and PESQ scores and lower LLR and WSS scores is expected to result in better speech intelligibility.

2.4.2. Technical Measures of Speech Intelligibility

Technical measurement of speech intelligibility uses either synthetic speech-like signals or parameters that include appropriate average speech and noise spectra. Since technical tests are fast and inexpensive, there are real-time and online applications that can be used instead of perceptual subjective tests. In fact, instead of measuring speech perception, they estimate the transmittance of the input signal. Technical tests are used more than perceptual tests, because they are more cost-effective and time-effective (Letowski, 2017).

The ANSI/ASA S3.5 and IEC 60268-16 standards recommend the use of three quantitative speech intelligibility predictors: the Articulation Index (AI), the Speech Transmission Index (STI), and the Speech Intelligibility Index (SII). In addition, the ITU recommends a group of algorithms for predicting speech quality, with the Perceptual Evaluation of Speech Quality (PESQ) algorithm being the most advanced. All of these measures use a psychoacoustic model to estimate the overall distortion introduced by a transmission system and calculate the resulting speech quality. According to Beerends et al. (2004; 2005; 2009) PESQ scores are generally found to be strongly correlated with speech intelligibility ratings, with correlations ranging from 0.91 to 0.99 under certain conditions.

2.4.3.1. SII

SII (Speech Intelligibility Index) is the most widely used quantitative speech intelligibility estimator standardized in ANSI/ASA. SII was created in 1997 as a major revision of the 1986 Articulation Index (AI), which is a measure developed by

communication engineers to quantify the relationship between the portion of the average speech spectrum that remains audible in the presence of filtering, noise distortion, and low speech level (Cahill, 2010). It is typically used in acoustical engineering to evaluate the performance of a space or system for speech communication.

The Speech Intelligibility Index Standard defines a method for calculating a physical measure that is closely related to the intelligibility of speech. It is based on acoustical measurements of speech and noise.

The SII model estimates the average speech intelligibility for a specific speech in a noisy environment for a group of people speaking and a group of people listening, regardless of gender. It is effective in difficult listening conditions, such as when there is background noise, filtering, or echoing. The SII is calculated using a formula that takes into account the level of speech in the environment and the level of background noise.

The SII is the percentage of total speech that can be heard by the listener. The standardized SII is a single value that summarizes the importance-weighted speech audibility across frequencies, including corrections for the effects of distortion, such as reverberation and masking. It can range from 0 to 1, where 0 indicates that no speech is audible and 1 means that all speech is audible.

The SII may be useful in both clinical and research settings, according to Scollie (2008). The SII has been used to estimate speech intelligibility for adults with normal hearing in realistic conditions and to predict the minimum level of speech needed to be understood in noisy conditions (Cahill, 2010). The SII has also been used to study hearing impairment and hearing aids as well as automated calculations with manufacturers' probe microphone equipment for children and adults.

In order to convert the AI (Articulation Index), SII (Speech Intelligibility Index), or STI (Speech Transmission Index) into a measure of speech intelligibility, a non-linear transformation must be applied. This transformation is based on empirical data and varies depending on the type of speech material being used. According to Brand (2008), the transformation is typically steeper (meaning there is a greater increase in intelligibility with each unit increase in the index) for materials that have context, such as sentences, compared to materials with single words.

There are four settings for calculating the SII:

1. Critical Frequency Band (21 bands),
2. One-Third Octave Frequency Band (18 bands),
3. Equally Contributing Critical Band (17 bands),
4. Octave Frequency Band (6 bands).

The basic steps for all four settings are the same, but the formulas differ in the number and size of the frequency bands. (Cahill, 2010)

The general formula for calculating the SII is:

$$\text{SII} = \sum_{i=0}^{n} I_i A_i \qquad (3)$$

The Speech Intelligibility Index (SII) is calculated by summing the product of the audibility of each frequency band (A_i) and its importance to speech understanding (I_i) across all frequency bands. The number of frequency bands used in the calculation is represented by n in the formula. The values of I_i, known as the frequency importance function (FIF), are based on specific speech stimuli and sum to approximately 1. The A_i values, which range from 0 to 1, indicate the proportion of audible speech cues in a given frequency band and are determined based on the level of speech relative to noise in that band. By multiplying A_i and I_i for each frequency band and summing the results, a single numerical index representing the overall speech intelligibility is obtained.

SII provides a metric between 0 and 1, where;

SII = 0: Unintelligible, perfect speech privacy

SII = 0.1: Confidential privacy

SII = 0.2: Normal privacy

SII = 0.3: 35% of sentences are correct

SII = 0.5: 90% of sentences are correct

SII = 1: Perfectly intelligible

According to the ANSI 1997 standard, a communication system is considered good if its SII is greater than 0.75 and poor if its SII is less than 0.45. (Cahill, 2010).

Is SII language dependent?

Studies have shown that the Speech Intelligibility Index (SII) is language-dependent. The SII is calculated as the proportion of speech information that is audible across a specific number of frequency bands. To determine the SII, the first step is to measure the audible dynamic ranges (DRs), which are defined as the range in decibels from the peak of a speech signal to the auditory threshold in specific frequency bands for people with hearing loss and for speech presented without added noise. The proportion of audible speech in each frequency region (band-audibility function, BAF) is then multiplied by the relative importance of that frequency region (band-importance

function, BIF). The resulting values are summed across all frequency bands, and the final result is a number between 0 and 1, with 1 indicating perfect intelligibility and 0 indicating that none of the speech was understood.

Components of the SII (BIF and BAF) can vary depending on the specific language being used. For example, the BIF may depend on the language due to differences in acoustic and linguistic characteristics (Jin, 2014, Jin et.al. 2017). Moreover, the differences in BIFs between languages can be attributed to differences in phonological characteristics, such as the tonal nature of a language (Wong et al. 2007). The dynamic range (DR) of speech, which is a factor in determining the BAF, may also vary across languages. These variations in linguistic and acoustic characteristics can lead to differences in the SII predictions for each language. Boothroyd et al. (1994) demonstrated that English phonemes have different frequency and level characteristics. In short, the SII can vary between languages because spoken languages have unique characteristics that affect the BIF and BAF (Jin, 2014.). Moreover, Chen (2016) found that different languages have distinct frequency band importance functions (FIFs), which are used to calculate the Speech Intelligibility Index (SII). This suggests that it is important to establish separate FIFs for different languages when using the SII. In addition, the study found that there are differences in the FIFs for male and female speech: (1) male speech was found to be more intelligible when using the full frequency band, while female speech was less intelligible; (2) the frequency band centered at 160 Hz was found to be more important for female speech than for male speech.

The frequency band importance function is a measure of how much each frequency band contributes to the understandability of speech. Each frequency band is assigned a value between 0 and 1, and the sum of all of these values is 1. The specific values assigned to each frequency band can vary based on factors such as the speaker's speech spectrum, the language being spoken, and the phonemic content of the stimuli (Damico, 2019)

Galburn and Kitapci (2016) studied how room acoustics affect the understandability of speech in four different languages (English, Polish, Arabic, and Mandarin). They found that there were significant differences between the languages in terms of both how well listeners could understand the speech and how well the results of predictive speech intelligibility tests predicted the understandability of the speech.

Studies related to SII

There are some studies related to SII. One of them was conducted by Jin, et al. (2017) in which they investigated the language sensitivity of the speech intelligibility index (SII) at 30 different dynamic ranges (DRs) in different languages. In the current version, a 30dB DR is used for all frequency ranges to calculate the SII. The researchers used DRs ranging from 10 to 60 dB for three different languages: English, Korean, and Mandarin. They found that the SII is not sensitive to the assumed dynamic

range (DR) and accurately predicts speech intelligibility in different languages and stimuli at different ranges of dynamic range.

Another study was done by Chen et al. (2016) In this study, they developed a Frequency Function for Mandarin Chinese using words spoken by men and women. Since there are differences between languages, they suggested developing separate frequency functions for different languages when calculating the SII. They also found that the frequency function was different for male and female speakers.

Nathwani et. al, (2017) used SII and PESQ objective tests when evaluating the performance of the speech modification model they developed in terms of intelligibility. They stated that the method with higher SII and PESQ scores has higher intelligibility.

Another study was done by Cahill (2010), who stated that SII is a helpful tool for patients and clinicians to accurately assess hearing loss. Scollie (2008) investigated whether SII could be used to predict consonant recognition scores of adults, children with and without hearing loss. Scollie (2008) indicated that there were significant differences in SII scores between these three groups and stated that this difference was probably due to hearing level and age. In addition, they emphasized that SII can be used both clinically and in research.

2.4.3.2. PESQ

The Perceptual Evaluation of Speech Quality (PESQ) algorithm is used to estimate subjective ratings of the quality of a degraded audio sample. It was developed using a large number of recordings of sentences spoken in various languages by different speakers, which were made using different speech encoders of varying quality and subjected to typical network transmission disturbances. (Köster, 2017). The algorithm returns a score from -0.5 to 4.5, with higher scores indicating better quality, based on listening tests in which a sufficient number of listeners rated the recordings on a speech quality scale from 1 (bad) to 5 (excellent). PESQ is able to analyze specific parameters of audio, such as time warping, variable delays, transcoding, and noise. Its ability to automatically evaluate the quality of any audio system makes it appealing for use in various applications. PESQ compares two perceptually transformed signals and uses the resulting noise disturbance value to estimate the perceived speech quality (Liu, 2006). It was standardized as ITU-T Recommendation P.862 in 2001, replacing the earlier PSQM algorithm (ITU-T Rec. P.861). PESQ has an improved time-alignment module that makes it more robust for use in real-world networks with varying delays.

PESQ is designed to measure the one-way quality of narrowband telephone signals and model the perceived speech quality in the listening context, which is primarily impacted by speech distortion from factors such as speech codecs, background noise, and packet loss (Guéguin, M. et al., 2008). It has been found to have a correlation of close to 0.935 with subjective data (J. G. Beerends et al., 2002, ITU-T Recommendation P.563, 2004). PESQ is able to evaluate listening speech quality

under various conditions of degradation, such as distortion from echo cancellers or noise reduction systems, packet loss, and distortion from codecs and ambient noise on the sending side. It detects speech frames in the degraded signal and extracts a set of parameters for the analysis of the vocal tract, the unnaturalness of speech, strong additional noise, interruptions, mutes, and time clipping. The goal of developing PESQ was to create a method for objectively measuring the quality of speech based on comparing the original, undegraded signal (the reference signal) with the degraded signal (the measured signal) and to ensure that this measurement correlates well with the results of the listening tests (Köster, 2017).

PESQ outputs quality indications that mimic the Mean of Opinion Score (MOS) and computes the final speech quality score using a linear combination of these different quality parameters (Beerends, 2002). According to Wei M. Liu (2008), the results indicate that there are good quality measures that could be used to estimate intelligibility. Moreover, studies have shown that the PESQ (Perceptual Evaluation of Speech Quality) has been found to be the most reliable objective measure for evaluating speech quality (Rix et. al., 2001, and Loizou, 2007), and to a certain extent, speech intelligibility (Hu et. al., 2007, Loizou, 2011).

PESQ score definitions are typically grouped into six bands.

1. 1.00- 1.99 No meaning understood with any feasible effort
2. 2.00 - 2.39 Considerable effort required
3. 2.40 - 2.79 Moderate effort required
4. 2.80 - 3.29 Attention necessary; a small amount of effort required
5. 3.30 - 3.79 Attention necessary; no appreciable effort required
6. 3.80 - 4.50 Complete relaxation possible; no effort required

According to Ali et al. (2009), PESQ results can be divided into two quality classes: the good quality part (scores > 2.5), and the bad quality part (scores < 2.5).

Is PESQ language dependent?

PESQ, or Perceptual Evaluation of Speech Quality, is a tool used to objectively measure the quality of speech. It was developed using a diverse set of recordings in multiple languages, and has been standardized by the ITU-T Rec. P.862 standard. PESQ is generally considered the most effective objective measure of speech quality, and its performance is typically evaluated by comparing its scores to subjective Mean Opinion Score (MOS) ratings. However, PESQ's performance has mainly been tested on English and some other European and Asian languages. A study by Ali et al. (2009) found that the accuracy of PESQ scores for Arabic speech may depend on the specific quality assessment model used, and that different languages can have different levels of stationarity due to their prosodic structure. This suggests that the effectiveness of PESQ and other objective quality assessment algorithms can vary depending on the language being evaluated.

2.5. Speech Perception

Speech perception is the process by which we understand and interpret spoken language. The goal of speech perception is to segment the continuous stream of sound into individual words and phonemes. It is a crucial aspect of human communication and allows us to understand and convey meaning through the spoken word. Speech perception is a complex and multifaceted process that involves a number of different cognitive and perceptual abilities, including the ability to recognize and distinguish different sounds, to understand the meaning of words and sentences, and to interpret the speaker's intention and emotional state (Kalikow, Stevens, & Elliott, 1977).

There are several factors that can impact how well a person understands spoken language, including their proficiency in the language being spoken, the speaker's accent, and the presence of background noise. For example, someone who is proficient in a language may have an easier time understanding speech in that language than someone who is not proficient. Similarly, a speaker with a strong accent may be more difficult for a listener to understand, especially if the listener is not familiar with the accent. Noise in the environment can also make it harder to understand speech, as it can mask or distort the words being spoken. In addition to these external factors, the clarity of the speech signal, the listener's level of attention, and their familiarity with the language being spoken can also affect speech perception. Some people may also have difficulties with speech perception due to hearing loss or language disorders like dyslexia.

To study the role of semantic context in speech perception, researchers have used a variety of methods, including perception tasks, eye gaze measurements, and brain imaging techniques. In perception tasks, listeners are presented with spoken words or sentences and asked to identify or repeat what they have heard. Eye gaze measurements can be used to track where listeners are looking while they listen to speech, providing insight into how they are processing the words and phrases being spoken.

Studies using these methods have consistently shown that semantic context can significantly improve speech perception. For example, research has demonstrated that listeners are more accurate at identifying words or phrases when they are heard in the context of a sentence or text, rather than in isolation (Miller, Heise, & Lichten, 1951). This is because listeners use the information, they have already heard to make predictions about what will come next, both within the sentence and in the wider context of the conversation (Drager & Reichle, 2001; Wingfield & Tun, 2007). Additionally, providing listeners with semantically related cues can improve their performance on speech perception tasks, especially in more difficult listening conditions (Zekveld, Rudner, et al., 2011). Eye gaze measurements using a "visual world" paradigm have also demonstrated the role of predictions in speech perception (Huettig, Rommers, & Meyer, 2011), with listeners focusing more quickly on objects associated with the action being described in a sentence (Altmann & Kamide, 1999).

The research on semantic context and speech perception clearly demonstrates the important role that semantic context plays in understanding spoken language. By using the meaning of words and phrases in relation to the wider context of the conversation or text, listeners are able to more accurately interpret and understand spoken language, even in challenging listening conditions.

In real-life situations, speech perception often occurs under non-ideal conditions such as the presence of background noise, reverberation, or channel distortion, which can make it difficult to understand spoken language. These types of conditions can often occur simultaneously, making it difficult to understand spoken language. For example, when speaking on the phone while walking on a busy street, there may be noise from the road as well as distortion caused by the phone connection. Most linguistic exchanges take place in the presence of other sound sources, and the ease of understanding language can be affected by factors such as the signal-to-noise ratio, the ability of the non-target noise to attract the listener's attention (e.g., competing speech may be more distracting than the noise from a washing machine), and the distance from the sound source (Mattys, Davis, Bradlow, & Scott, 2012).

2.6. Measurement Methods for Listening Attention

There is no standardized test for measuring listening attention. There are several non-standardized approaches to measure auditory attention, including:

- Event-related potentials (ERPs): This technique involves measuring the electrical activity of the brain in response to specific auditory stimuli (Michie et al., 1990; Schröger, 1996; Geal-Dor et al., 2005).
- Eye tracking: This method measures eye movements and gaze patterns in response to auditory stimuli.
- Pupillometry: This technique measures changes in pupil size in response to auditory stimuli, which is an indirect measure of attentional effort.
- Speech reception threshold (SRT): This method involves measuring the minimum level of sound that an individual can detect, which provides an indication of the person's ability to attend to auditory stimuli.
- Dual-task paradigm: This technique involves measuring performance on a primary task while also presenting auditory stimuli, with the aim of assessing the impact of the stimuli on task performance.
- Auditory change detection tasks: This method involves measuring the ability to detect changes in auditory stimuli, which is an indicator of attentional allocation to the auditory modality (Schröger, 1996; Berti, 2012; Pannese et. al., 2015).
- Functional magnetic resonance imaging (fMRI): This technique measures brain activity by detecting changes in blood flow, which can provide information about the areas of the brain that are engaged during auditory attention.

We explained eye tracker, pupillometry, pupil size, and dual task paradigm methodology in detail.

2.6.1. Eye Tracking and Pupillometry

Van Engen et al. (2018) stated that eye tracking provides precise, moment-to-moment data on the locus of visual attention by recording the direction of listeners' eye gaze. This allows researchers to evaluate the progression of target word recognition and the different interpretations listeners may entertain during recognition. On the other hand, pupillometry measures pupil dilation, a physiological indicator of cognitive load, thereby providing a quantitative measure of mental effort during speech recognition without relying on behavioral responses.

McMurray and colleagues (2017) employed eye tracking to examine lexical activation and competition in listeners with cochlear implants (CI) and normal hearing (NH) listeners. The participants were presented with six-word pairs, such as "beach" to "peach" and "sip" to "ship." They were asked to select the picture corresponding to the word they heard, and their eye movements were monitored to investigate their consideration of the other word in the pair.

Brouwer et al. (2012) found that competition depended on the speech style of the speaker, whether reduced or fully articulated. When the words were reduced in the sentence around a target, listeners looked at competitors that were similar to the canonical form. However, when the surrounding speech was fully articulated, listeners showed a clear preference for the canonical competitor. Brouwer et al. (2013) showed that for reduced form recognition, the supporting discourse context was more important than the canonical form.

Winn et al. (2018) noted that each measure of effort has its advantages and limitations. Reaction times can be influenced by the subject's dexterity, speed, age, and physical ability. Similarly, pupil size and range of dilation can be affected by age. Winn et al. (2018) also stated that while pupillometry is not without limitations, it is simple and fast to set up.

Recent studies have combined eye tracking and pupillometry to simultaneously examine listeners' attention direction and cognitive load during speech processing. Van Engen et al. (2018) stated that combining these two methods can provide insight into the relationship between cognitive effort and changes in lexical activation and competition dynamics, but several methodological challenges must be addressed to achieve meaningful results.

2.6.1.1. Eye Tracker Methodology

In eye tracking methodology, a researcher observes and measures individuals' eye movements to determine where they are looking at a given time. The researcher also examines eye movements on a screen. To perform this technique, an infrared camera should be placed near or under the display monitor on a uniform desktop computer. This allows the camera to identify and record even the slightest eye movements and their characteristics. This is mainly achieved through the infrared light from the

camera's LED, which creates strong reflections in the eye's features, making them easy to track.

There are two basic types of eye movements in eye-tracking research: fixations and saccades. Fixations are defined as moments when the eyes are relatively stationary and encoding information (Poole & Ball, 2005). In other words, the user fixes their gaze on an object on the screen (Nielsen & Pernice, 2010). Information is visually extracted during fixations because the eyes are then relatively still and focused. On the other hand, a saccade is an eye movement between fixations, lasting 20 to 35 milliseconds (Poole & Ball, 2005). It can also be defined as a quick movement of the eyes from one fixation to the next (Nielsen & Pernice, 2010).

2.6.1.2. Pupil Size

The pupil is the center of the colored part of the eye and controls the amount of light entering the eye. In bright light, the pupil gets smaller, and in dark light, it gets larger. Moreover, pupil size is used to measure emotional arousal in the literature. In addition, it can also be used to measure mental workload. For instance, pupils generally enlarge with increased processing demands.

There are two types of pupil methods in the literature: the Bright-Pupil method and the Dark-Pupil method. Both methods aim to create maximum contrast between the pupil and iris to make the pupil easily recognizable (Bojko, 2013). However, the effectiveness of each method depends on the participant's eye characteristics, such as eye color. The Bright-Pupil method works better in dark environments and with bright-colored (blue) eyes, while the Dark-Pupil method works better in regular indoor lighting conditions and with dark-colored (brown) eyes (Bojko, 2013). In our study, we will use the Dark-Pupil method to reach more participants.

Cognition-related processes, such as attention, are linked to pupil reaction patterns (Beatty, 1982). Larger pupil dilation occurs in tasks with higher cognitive demands. In general, pupil dilation increases with increased processing load during task performance. Listening-related effort and fatigue can also be characterized by changes in physiological arousal. Pupillometry, a technique based on measuring pupil size, can be used to assess altered physiological arousal and attention in real-time while performing cognitive tasks such as a listening exercise (Koelewijn et al., 2015).

Several studies in the literature have used pupil size to measure listening effort. For example, Koelewijn et al. (2015, 2017) used pupil response methods to measure listening effort in difficult attention conditions and to show the effects of attention on speech reception threshold in normal and impaired hearing individuals. McGarrigle et al. (2017) measured listening-related effort and fatigue in school-aged children using pupillometry to investigate the effect of signal-to-noise ratio on listening effort and fatigue.

Winn found that pupil size was a more sensitive measure than the dual-task paradigm. Winn suggested designing stimuli that are not too difficult or too easy, and provided recommendations for the experiment design, including waiting time, number of trials, timing, and lab conditions. They recommended waiting 4 to 6 seconds after each

participant's response to allow pupil size to return to baseline levels and having at least 16 to 18 good recordings of pupil size for each condition. They stated that the timing of each event should be carefully planned.

Recognizing pupil size poses several challenges. For instance, wearing glasses can obstruct pupil detection, and mascara can confuse the pupil with mascara-covered lashes (Bojko, 2013). Winn et al. (2018) suggested avoiding the use of bright colors, particularly white backgrounds and computer screens, as they can present multiple problems. For example, excessive pupil constriction and discomfort for participants, potentially requiring additional breaks. They advised taking the following factors into consideration when choosing participants: eye color, makeup, age, hearing status, pharmacological effects, caffeine intake, eye diseases, history of head injury or neurological problems, general hearing ability, native language, and fatigue. Zekveld et al. (2018) listed the factors that affect pupil size during auditing and provided the number of publications that demonstrate the impact of each factor on pupil size.

Factors Associated with Input:

- Demands Linked to the Input
- Source Characteristics
- Complexity of Auditory Processing
- Complexity of Linguistic Elements
- Higher Memory Load
- Demands of Memory Processing
- More Complex Cognitive Challenges
- Word Meanings

Factors Related to Transmission:

- Level of Degradation
- Type of Degradation

Listener-Related Factors:

- Age
- Proficiency in Non-Native Languages
- Hearing Abilities
- Experience with Auditory Training
- Cognitive Capabilities
- Musical Expertise
- Gender
- Fatigue

Automatic Attention Related Factors:

- Presentation of Auditory Stimuli
- Sound Level

Intentional Attention Related Factors

- Increased Attentional Requirements

Factors Associated with Motivation and Arousal

- Motivational Reward Mechanisms
    - Emotional Valence and Engagement
    - Deception
    - Assessment of Capacity Demands

Factors like Fatigue, Arousal, and Displeasure

- Time Spent on the Task

2.6.1.3. Pupillometry

Pupillometry is the study of how the size of the pupil (the black circle in the center of the eye) changes, and it is often measured using electronic devices such as eye trackers. This field of research was originally called pupillography, because it involved creating a graphical trace of pupil size changes. However, as technology advanced, the term pupillometry replaced pupillography, because the "graphic" aspect was no longer emphasized.

In 1960, Hess and Polt demonstrated that pupil size increased when adults viewed images of half-naked members of the opposite sex and when female participants saw pictures of babies. This study helped establish pupillometry as a valid method of study in the scientific and psychological communities. In the years since, it has been shown that pupil size can also change due to other forms of cognitive processing, such as working memory load including remembering strings of digits (Beatty & Kahneman, 1966), performing difficult mathematical calculations (Landgraf, Van der Meer, & Krueger, 2010) and attention to speech in difficult listening conditions (Zekveld et al., 2010).

Pupil Outcome Measures

In the literature on pupillometry, there are four main pupil outcome measures that are typically considered:

- pupil baseline (the average pupil diameter in the time before the stimulus starts (generally 1 or 2 seconds),
- mean pupil dilation relative to the baseline (calculated over a variable time window),
- peak pupil dilation (the maximum positive deviation from the baseline over a variable time window),
- latency of the peak dilation (the time it takes for the peak dilation to occur).

The interpretation of the main pupil outcome measures depends on the experimental stimuli and design. Mean pupil dilation is generally considered a more reliable measure of cognitive resource allocation than peak pupil dilation (Ahern & Beatty, 1979; Verney, Granholm, & Dionisio, 2001). In certain cases, the pupil response can be analyzed using time-series methods to evaluate cognitive effort over time within selected time windows (Winn et al., 2018).

There are two main methods for baseline correction in the literature: proportional (or divisive) baseline correction (corrected pupil size = pupil size/baseline) and subtractive baseline correction (corrected pupil size = pupil size – baseline). The subtractive method will be used in the study presented in this thesis because it is less affected by data distortion, such as blinks or data loss (Mathôt et al., 2018). Moreover, the subtractive method facilitates straightforward comparisons between experiments, ensuring normalization of pupil size changes from 0 mm. This comparison holds regardless of individual anatomical distinctions or disparities in luminance levels across different laboratory settings (Reilly et al., 2018). The pupil response for the studies will be presented in millimeters rather than as a percentage change.

2.6.1.4. Pupillometry in Language and Listening Research

Measuring changes in pupil size can be a helpful tool in cognitive research because it is linked to central cognitive processing (Beatty & Lucero-Wagoner, 2000). It is inexpensive and non-invasive, and provides real-time insights into brain activity. Measuring pupil size, along with eye movements, can give researchers real-time insights into how the brain is processing information. It is also useful for research with subjects who are unable to communicate or cooperate (Laeng et al., 2012).

Pupillometry has several characteristics that make it a useful tool for studying listening studies. It is a time-series measurement, which means that multiple measures of pupil size are taken during the task. This is important because listening effort involves both quick auditory encoding and ongoing cognitive processing over time. Pupillometry is also suitable for testing individuals who use assistive devices like hearing aids and cochlear implants, because it does not interfere with these devices like some other measurement techniques.

Pupillometry has been used to study the effects of task characteristics on pupil size in language processing research. It has been applied to a range of listening tasks that vary in complexity (Kramer et al., 2013), and performance is typically measured by the number of correctly understood sentences or words. Speech intelligibility can be manipulated by adding masking noise or degrading the quality of the auditory stimuli.

In a study by Zekveld et al. (2010), speech intelligibility was controlled by adjusting the signal-to-noise ratio (SNR) to different levels and asking participants to perform speech reception threshold tests. When listeners heard sentences in noise, the size of their pupils increased and the time it took for the pupils to reach their maximum size

increased as the intelligibility of the speech decreased, suggesting that it took more cognitive effort to process speech with lower intelligibility.

Pupil dilation has been shown to be correlated with the intelligibility of spoken language, with the maximum dilation occurring at intermediate levels of intelligibility. However, the relationship between pupil dilation and intelligibility is more complex than a simple linear relationship (Zekveld & Kramer, 2014), and can be influenced by other factors such as individual differences in the ability to comprehend masked sentences and the level of task engagement. At very low levels of intelligibility, pupil dilation may be smaller (Ohlenforst et al., 2017; Zekveld & Kramer, 2014). It can be due to a tendency to give up or due to the impact of low motivation on pupil response. This may also be related to the FUEL model, by the evaluation of demands on capacity and the allocation of resources (Pichora-Fuller et al., 2016).

In 2012, Koelewijn et al. conducted a study in which they measured pupil dilation in participants as they listened to spoken sentences in different types of noise. The results showed that the pupils of the participants had a larger response to the single-talker masker compared to the other types of noise, although the participants performed slightly better on the task when the single-talker masker was present.

There is a significant amount of research on the relationship between pupil dilation and the linguistic complexity of spoken stimuli. It has been found that pupil dilation is affected by factors such as semantic difficulty, word frequency, lexical competition, semantic processing, and stimulus neighborhood density (Kuchinsky et al., 2013; Kuipers & Thierry, 2011; Schmidtke, 2014). For example, larger pupil responses occur when attending to semantically difficult words and low-frequency words particularly in high memory-load conditions (Chapman & Hallowell, 2015; Elshtain & Schaefer, 1968). Pupil dilation is also related to learning outcomes for unknown words and can be influenced by syntactic complexity, sentence ambiguity, and prosody (Papesh, Goldinger, & Hout, 2012). Pupil responses have also been found to increase with the syntactic complexity of sentences, and are more strongly correlated with grammar complexity than with subjective comprehension (Schluroff, 1982). The pupil response is also affected by sentence complexity.

2.6.2. Dual Task Paradigm

We will employ a dual-task paradigm to assess behavioral selective directional auditory attention. The primary task in this study will be the recognition of monosyllable words, while the secondary task will be the recall of three words among those presented in the primary task.

Ohlenforst (2017) reviewed studies that used the dual-task paradigm, stating that the primary task usually involves word or sentence recognition, while the secondary task can range from response time tasks, memory tasks, tactile pattern recognition tasks, or driving simulation. Sahlén et al. (2018) used this paradigm in children by having them match recorded sentences to one of four pictures. The children listened to 80 sentences

and on each trial, matched the sentence to a picture with a mouse click. Pals et al. (2019) used a visual rhyme judgment task as the secondary task in their study, where participants were presented with a pair of monosyllabic Dutch words and asked to determine if they rhymed. Participants responded by pressing a key, and their responses and reaction times were recorded by software. The words remained on the screen until a response was made or until a set time elapsed. Gagné et al. (2017) listed compilation of secondary task varieties employed in examined dual-task investigations on speech listening effort.

- Tracking a visual motor task exhibited on a computer screen
- Measuring response time to a visual probe or distractor
- Recollection of words presented either prior to or during the primary task
- Assessment related to an aspect of the primary task
- Recognition of tactile patterns
- Engagement in a semantic judgment activity
- Operating a car simulation
- Mentally rotating Japanese characters
- Determining response time to specified auditory signals
- Making rhyme judgments for visually presented Word

CHAPTER 3

METHODS AND EXPERIMENT DESIGN

In this section, we will present our proposed approach. In this study, to measure selective auditory attention and its relation to the speech intelligibility measures of SII, PESQ and SNR, we used spoken three-letter monosyllabic Turkish words as stimuli with eye–tracking and pupil dilation measurements under different noisy conditions.

3.1. Speech Material

The speech materials consisted of three-character monosyllabic Turkish words. We used a blog to reach new Turkish words published by Şule Kişioğlu, who is an audiologist and and a physiotherapist (Kişioğlu, 2013). Also, we searched (Kelime Listelerinde Arama, 2023) for each letter (Figure 1). For example, we chose the option “the word beginning with letter b” in the database and listed all words that begin with the letter “b”. Then we selected the three-character monosyllabic words for each letter. Also, due to nature of the Turkish language, there were no words begin with "Ğ" letter.

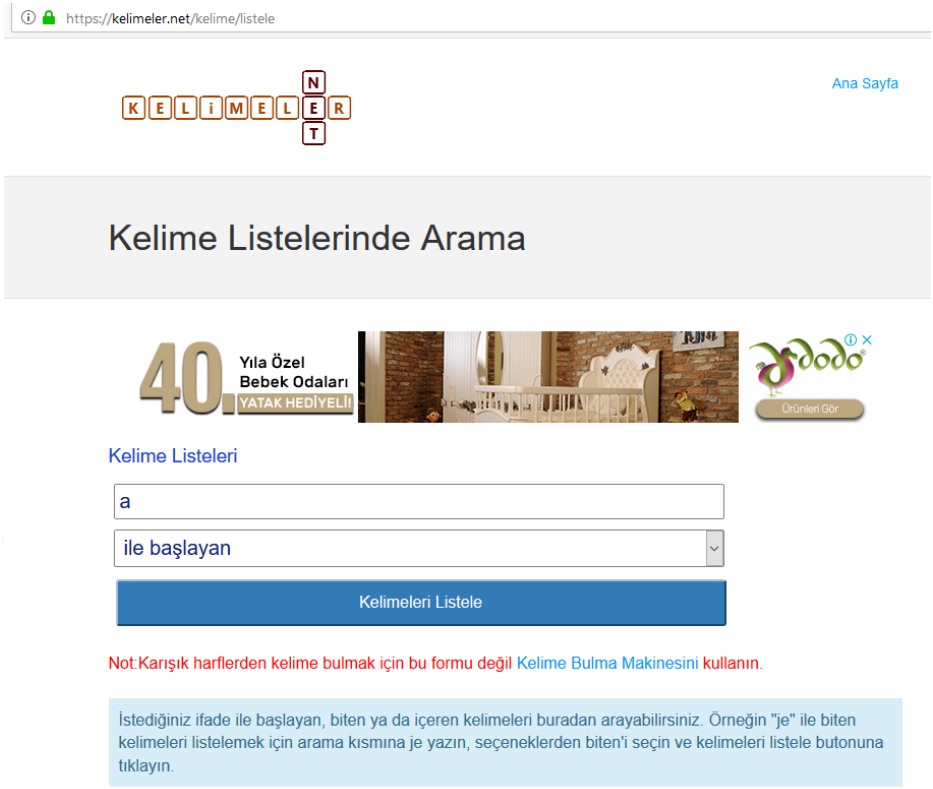


Figure 1: Searching three-character monosyllabic Turkish words in kelimeler.net

3.2. Recording of Speech Material

The speech materials consisted of three-character monosyllabic Turkish words. The speech materials were vocalized by two professional narrators with experience on recordings and diction training. One of them was a 28-year-old female and the other was a 31-year-old male. They were both native Turkish speakers.

The words were spoken one by one by the narrator with silence in between and recorded with 44.1 kHz sampling frequency. Then, each word was saved as a separate wave file, synchronizing the starting time of each word using the Audacity software. Later with the help of MATLAB, the length of each audio file was fixed as 1 second-long and 1second-long silence was added at the end of each word, resulting in 2 s duration for each word.

All speech materials were recorded in the SPARG (Spatial Audio Research Group) Lab at Modeling and Simulation Center (MODSIMMER) as shown as Figure 2. The laboratory housed an 11 m^2, soundproof room, whose background noise was less than 40 dB SPL. The words were recorded by a MacBook Pro (2011) laptop running Audacity software with Focusrite Scarlett 18i8 USB audio interface sound card and BM-800 (cardioid) vocal microphone as shown in Figure 2.



Figure 2: METU SPARG (Spatial Audio Research Group) Lab at the Modeling and Simulation Center

3.3. Stimuli

In each stimulus, ten different three-character monosyllabic Turkish words were present, which were prepared by selecting and combining the edited 2 s-long word recordings. The resulting duration of each stimulus was 20 s. Each word was used only once in the stimuli. For example, “tam, sek, ker, mil, gel, yak, sos, sol, şad, poz” was one of the stimulus in which all words are different.

The stimuli were listened to by the participants using closed-cup professional headphones. The presented stimuli sound level was the same throughout the listening tests. Depending on the test conditions, the stimuli voiced by the female or male speaker were played alone, i.e., without interference, to the left ear alone, right ear alone or bilaterally; or together with another stimuli to the left ear or right ear, while the interference voiced by the other gender was played to the other ear. The information on which speaker would be played to which ear was not shared with the participants. The participants were informed about the speaker gender (female/male) that they should focus on.

Twelve different listening conditions were created; covering two speakers (Female and Male), three speaker locations (Left, Right, Bilateral) and two interference cases (with and without interference). Each condition was tested three times, resulting in 36 trials as shown in Table 1. Trials were presented to the participants in random order, separated by short breaks. At the end of each trial, the participants were asked to write 3 words voiced by the target speaker.

Table 1: Stimuli Presentation

Condition	**Target Speaker Gender**	**Location of Target Source**	**Competing Speaker Gender**	**Location of Competing Source**	Number of Trials
1	Female	Right	Male	Left	3
2	Female	Left	Male	Right	3
3	Male	Right	Female	Left	3
4	Male	Left	Female	Right	3
5	Female	Right	-	-	3
6	Female	Left	-	-	3
7	Male	Right	-	-	3
8	Male	Left	-	-	3
9	Female	Bilateral	Male	Bilateral	3
10	Male	Bilateral	Female	Bilateral	3
11	Female	Bilateral	-	-	3
12	Male	Bilateral	-	-	3

Due to the limited number of three-character Turkish words, only 20 different word groups could be generated. Each word group, except four of them were played as target twice; with and without interference. Table 2 shows the presented stimulus which contains 10 Turkish three-character monosyllabic word.

Table 2: Presented stimulus which contains 10 Turkish three-character monosyllabic word

Trial ID	**Target Speaker Gender**	**Target Speaker Location**	**Competing Speaker Gender**	**Competing Speaker Location**	**The stimuli voiced by Female Speaker**	**The stimuli voiced by Male Speaker**
1	Female	Left	Male	Right	"tam","sek","ker","mil","gel","yak","sos","sol","şad","poz"	"nam","kil","tav","dur","bel","kel","kek","tay","gol","men"
2	Female	Right	Male	Left		"tas","boz","bek","ben","bay","hal","kez","bar","baz","kem"
3	Female	Bilateral	Male	Bilateral	"san","nur","bal","sim","toz","kın","pek","bas","saz","sit"	"saf","gam","sur","fol","pin","çam","kır","şak","kış","tek"
4	Female	Left			"her","kır","zam","hat	

Trial ID	Target Speaker Gender	Target Speaker Location	Competing Speaker Gender	Competing Speaker Location	The stimuli voiced by Female Speaker	The stimuli voiced by Male Speaker
					","sap","kış","yol","çan","kan","fen"	
5	Female	Right			"çil","gol","dek","kak","yas","dil","nam","tak","kam","tim"	
6	Female	Bilateral			"naz","fan","hap","hin","zil","tek","ten","bit","vur","put"	
7	Male	Left				"yay","taş","faz","sop","koz","dam","sos","kaş","kak","ver"
8	Male	Right				"sac","tim","doz","kur","yaş","bek","çay","yat","san","ker"
9	Male	Bilateral				"yaz","yel","naz","hak","pek","lal","han","kaç","bor","zam"
10	Male	Left	Female	Right	"zor","ban","gür","kap","gün","hol","hem","tur","gaz","şap"	"laz","tak","şan","yem","kas","yoz","dil","kut","jel","güm"
11	Male	Right	Female	Left	"gem","lor","caz","dut","dal","sac","şer","men","sur","yen"	"kıl","fas","zat","kız","kim","has","tut","bin","mor","put"
12	Male	Bilateral	Female	Bilateral	"nem","doz","çit","çin","cam","şan","bol","kol","yan","fal"	"gün","pim","tat","sit","kep","ham","şer","poz","yağ","yar"
13	Female	Left	Male	Right	"çil","gol","dek","kak","yas","dil","nam","tak","kam","tim"	"gaz","sal","mat","çim","lor","gül","son","var","kaz","fen"
14	Female	Right	Male	Left	"naz","fan","hap","hin	"laz","tak","şan","yem

Trial ID	Target Speaker Gender	Target Speaker Location	Competing Speaker Gender	Competing Speaker Location	The stimuli voiced by Female Speaker	The stimuli voiced by Male Speaker
					","zil","tek","ten","bit","vur","put"	","kas","yoz","dil","kut","jel","güm"
15	Female	Bilateral	Male	Bilateral	"her","kır","zam","hat","sap","kış","yol","çan","kan","fen"	"yay","taş","faz","sop","koz","dam","sos","kaş","kak","ver"
16	Female	Left			"san","nur","bal","sim","toz","kın","pek","bas","saz","sit"	
17	Female	Right			"nem","doz","çit","çin","cam","şan","bol","kol","yan","fal"	
18	Female	Bilateral			"hit","keş","cin","hay","zar","baş","haz","kaş","güz","bul"	
19	Male	Left				"nam","kil","tav","dur","bel","kel","kek","tay","gol","men"
20	Male	Right				"saf","gam","sur","fol","pin","çam","kır","şak","kış","tek"
21	Male	Bilateral				"tas","boz","bek","ben","bay","hal","kez","bar","baz","kem"
22	Male	Left	Female	Right	"gem","lor","caz","dut","dal","sac","şer","men","sur","yen"	"yay","taş","faz","sop","koz","dam","sos","kaş","kak","ver"
23	Male	Right	Female	Left	"dur","gam","kıl","pak","vay","sen","şak","han","jel","sop"	"sac","tim","doz","kur","yaş","bek","çay","yat","san","ker"
24	Male	Bilateral	Female	Bilateral	"tam","sek","ker","mi	"gaz","sal","mat","çim

Trial ID	Target Speaker Gender	Target Speaker Location	Competing Speaker Gender	Competing Speaker Location	The stimuli voiced by Female Speaker	The stimuli voiced by Male Speaker
					l","gel","yak","sos","sol","şad","poz"	","lor","gül","son","var","kaz","fen"
25	Female	Left	Male	Right	"hit","keş","cin","hay","zar","baş","haz","kaş","güz","bul"	"kıl","fas","zat","kız","kim","has","tut","bin","mor","put"
26	Female	Right	Male	Left	"gem","lor","caz","dut","dal","sac","şer","men","sur","yen"	"yaz","yel","naz","hak","pek","lal","han","kaç","bor","zam"
27	Female	Bilateral	Male	Bilateral	"zor","ban","gür","kap","gün","hol","hem","tur","gaz","şap"	"sac","tim","doz","kur","yaş","bek","çay","yat","san","ker"
28	Female	Left			"dur","gam","kıl","pak","vay","sen","şak","han","jel","sop"	
29	Female	Right			"tam","sek","ker","mil","gel","yak","sos","sol","şad","poz"	
30	Female	Bilateral			"gem","lor","caz","dut","dal","sac","şer","men","sur","yen"	
31	Male	Left				"gün","pim","tat","sit","kep","ham","şer","poz","yağ","yar"
32	Male	Right				"gaz","sal","mat","çim","lor","gül","son","var","kaz","fen"
33	Male	Bilateral				"laz","tak","şan","yem","kas","yoz","dil","kut","jel","güm"
34	Male	Left	Female	Right	"her","kır","zam","hat	"nam","kil","tav","dur

Trial ID	Target Speaker Gender	Target Speaker Location	Competing Speaker Gender	Competing Speaker Location	The stimuli voiced by Female Speaker	The stimuli voiced by Male Speaker
					","sap","kış","yol","çan","kan","fen"	","bel","kel","kek","tay","gol","men"
35	Male	Right	Female	Left	"naz","fan","hap","hin","zil","tek","ten","bit","vur","put"	"saf","gam","sur","fol","pin","çam","kır","şak","kış","tek"
36	Male	Bilateral	Female	Bilateral	"çil","gol","dek","kak","yas","dil","nam","tak","kam","tim"	"tas","boz","bek","ben","bay","hal","kez","bar","baz","kem"

3.4. Participants

In this study, we mainly focus on measuring selective auditory attention among normal-hearing young people. Tests were carried out with 38 participants; 19 female and 19 male (M_{age}=27.31), most of whom were graduate-level students. They all had normal or corrected-by-lenses vision and with no self-reported hearing problems or history of neurological diseases. All were native Turkish speakers and provided written consent in accordance with the Ethics Committee of the Middle East Technical University (METU). Only one of the participants was left-handed. Participants were not aware of the hypotheses being tested. The subjects were informed about the flow of the experiments and the measurement of their eye movements. They were asked to behave as they would in everyday life. Before the experiments, they were tested with an online hearing test. As stated in the literature, the participants of the listening studies are required to pass a hearing screening at 20 dB HL at octave frequencies from 250 to 8000 Hz. bilaterally.

3.5. Apparatus

The experiment was performed in FLE Eye Tracking Lab in METU as shown in Figure 3. The experiment was conducted on a PC with a 22-inch LCD (1920x1080 pixels; refresh rate: 60 Hz). The stimulus presentation and the recording of keyword responses were controlled by the Experiment Builder (SR Research Experiment Builder, Inc., Mississauga, Ontario, Canada), which is a graphical programming environment for running experiments. All stimuli were designed via Experiment Builder with 1920x1080 pixel resolution. The presentation of stimuli was executed through the utilization of the EyeLink 1000 Tower (SR Research, Inc., Mississauga, Ontario, Canada), a robust apparatus known for its proficiency in obtaining accurate measurements of pupil size. Pupil diameter measurements were acquired by means of the SR Research EyeLink system, employing a sampling frequency of 500 Hz within the monocular tracking mode. Also, we recorded eye and eye gaze data to observe and

measure the eye movements of individuals to find out where an individual is looking at a given time.



Figure 3: Test environment

3.5.1. EyeLink 1000 Tower

This section provides information about the EyeLink 1000 eye tracker. With its strong technical features and capacity, EyeLink can be used in many research areas. Camera technology and software support all configurations and provide solutions for eye and gaze tracking. EyeLink 1000 has two high-speed cameras with Standard and Fiber optic features and these cameras are designed to be used in different mounting options as shown as Figure 4. With the option of head stabilization, recordings that require high sensitivity can be made.

The system consists of 2 different computers. The main computer is dedicated to data collection, and the other computer is used as the Display Computer, where stimuli are presented to the participants. These two computers are connected to each other with an Ethernet link and information is shared between them.

In our study, stimuli presentation was made with EyeLink 1000 Tower. As the EyeLink 1000 is used in eye and gaze monitoring, it helps to get reliable pupil size measurements. Pupil diameter was recorded using a 500 Hz sampling frequency in monocular tracking mode. Using eye and gaze data, it can be observed and measured where an individual is looking at a given time. It can also be determined how the individual's eyes change from one position to another.



Figure 4: EyeLink 1000 Tower (Figure is taken from EyeLink 1000 User Manual)

3.5.2. Software: Experiment Builder

All stimuli were designed via Experiment Builder with 1920x1080 pixel resolution. Several third-party software support EyeLink1000. However, we chose Experiment Builder to create our study due to the extensive number of examples and video tutorials to learn and use the software.

SR Research Experiment Builder is software for creating visual experiments as shown as Figure 5. It is highly preferred by psychologists and neuroscientists as it provides ease of use without requiring any programming knowledge. It provides seamless integration with the SR Research EyeLink eye tracker device. Data collected from experiments prepared with Experiment Builder can be integrated into EyeLink Data Viewer and data analysis can be done easily with Data Viewer.

Experiment builder is a useful software for eye tracker studies. Camera setup, calibration, validation, and drift correction actions used in eye tracker studies are supported. Participants have the capability to design experiments wherein each trial displays a stationary screen featuring text or images, subsequently pausing to await participant responses. Its interface is as shown in the picture below. While designing an experiment, you can drag and drop the components you will use in your experiment. Components in the workspace perform the experiment in a certain flow. You can change and control the properties of each component. In addition, you can send a message to EyeLink during the action moments of some components and use these message moments in data analysis. You can add the data source (sound, text, image etc.) that you will use in the experiment to the library and present it to the participants randomly.

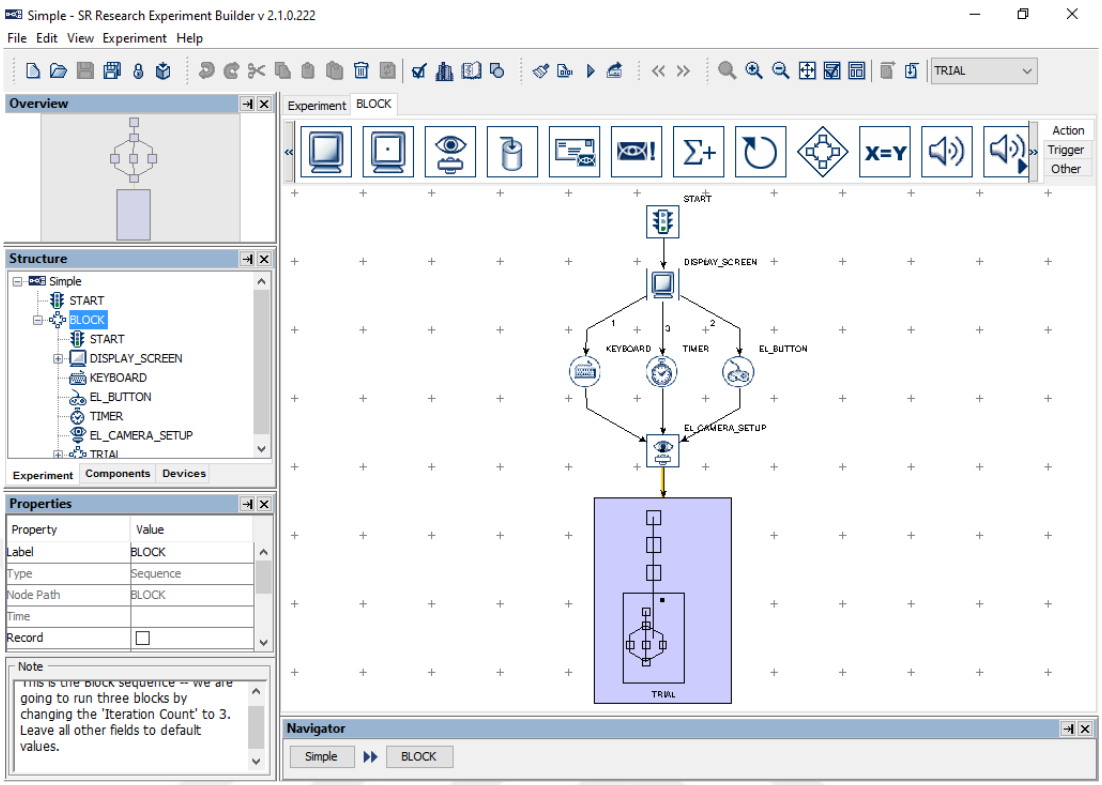


Figure 5: Experiment Builder

3.6. Experiment Design

Figures 6 to 10 display the screens shown to the participants during the experiment.

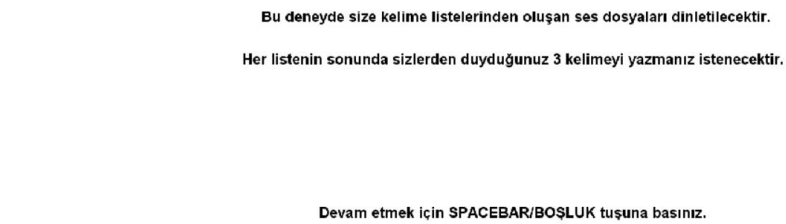


Figure 6: View of the instruction screen shown to participants

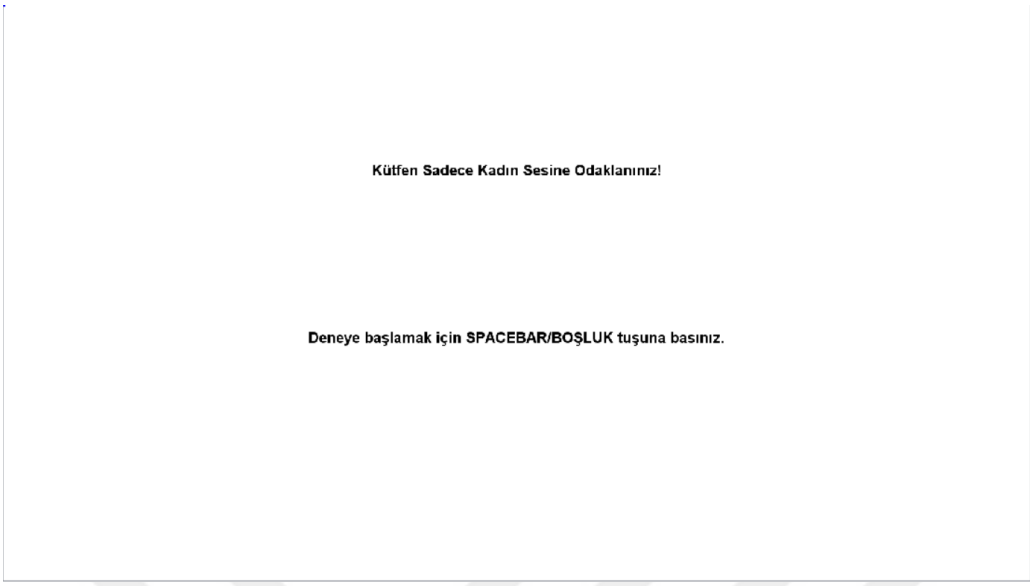


Figure 7: View of the attention clue given to the participants

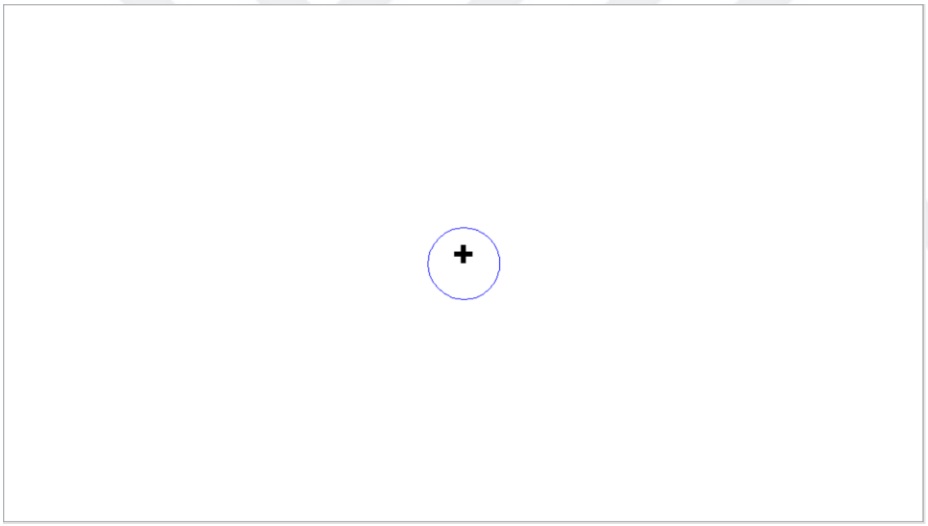


Figure 8: Displayed marker before playing the audio



Figure 9: Blank screen displayed while the audio is played



Figure 10: View of the screen where participants type the words they recall

3.7. Task and Procedure

At the beginning of the experimental session, each participant filled out a survey consisting of six questions about their gender, age, educational background, handedness, and eyes and ears' health. Then, each participant was tested alone in a dimly lit environment with their heads stabilized on a chinrest, while wearing wired on-ear headphones (JBL Tune 500). The signals can be conveyed through either speaker in a free field setting or headphones. Although the free field condition is more realistic, it is more challenging to carry out experiments, difficult to calibrate and slight listener movements can impact results. On the other hand, using headphones has benefits such as improved reproducibility for each individual listener, and easier setup, but requires more complex individual calibrations and is not suitable to use with hearing aids (Brand, 2008).

Participants were tested in a dimly lit environment with their heads stabilized on a chinrest. The chin and head alignment were explicitly adjusted for each participant, and they were put in a position where they could feel comfortable. Participants were seated at a distance of approximately 60 cm from the display screen as shown in Figure 11.



Figure 11: A picture from Experiment

At the beginning of the experimental session, each participant was instructed to fix their gaze on a small red cross displayed on the computer monitor with a gray background. A 9-point (3 x 3 square) calibration and validation were applied at the beginning of each experimental session. At the beginning of each trial, an initial task cue denoting the target speaker (female or male) for the participant to focus on was centrally presented on the screen. Subsequently, a fixation cross, measuring 39 x 116 pixels, emerged at the screen's center against a gray backdrop, remaining visible for a duration of 2 seconds. The participants directed their gaze towards this fixation cross, thereby ensuring consistent eye alignment at the trial's outset. Following the fixation

cross's disappearance, the stimulus was initiated, synchronized with the sound's commencement. The auditory stimuli played while the screen remained a uniform gray background, maintaining this configuration for a duration of 20 seconds per trial. Moreover, an additional 100 ms delay ensued, during which the screen remained devoid of content. While the participants' gaze was still being recorded, the response page was displayed, and the participants wrote three words they listened to using a keyboard. After writing each word, they pressed "Enter." To finish writing or the trial, they pressed "End." After every trial, to assess the necessity for recalibration, a black dot materialized at the screen's center against a white backdrop. The participant's task involved maintaining fixation on this black dot for a duration exceeding 1 s, during which the experimenter ensured accurate fixation detection before proceeding with the experiment. Otherwise, the calibration process was repeated. The upcoming trial started immediately after the presentation of the fixation cross. The entire session took about 20 minutes per participant.

3.8. Data Analysis

The spatial and temporal characteristics of eye movements were derived through the utilization of the EyeLink Data Viewer (SR Research Ltd., Mississauga, Canada). The analysis involved scrutinizing the horizontal and vertical coordinates of eye position over the course of time.

Data Viewer is a powerful analysis tool that allows users to display, filter, and report output of EyeLink 1000. Data collected from EyeLink 1000 is fully integrated with it. Data Viewer is fully integrated with eye-tracking data files that were collected using Experiment Builder. Data Viewer can read the custom messages created by Experiment Builder. These messages are critical for creating interest periods in the analyses part. Data viewer has powerful and flexible options for data management, processing, visualization, and output. It can help researchers to determine everything from whether participants looked longer at an object at the screen compared to others. Data Viewer has many different report options, allowing to create time series reports, output pupillometry data, fixation reports, saccade reports, trial reports and generate useful graphics such as heat maps. Each report has dozens of possible variables, allowing users to choose from depending on the research demands. We generated, trial reports, fixation, and saccade reports for each trial to analyze.

Data Viewer makes the analysis of eye-tracking data more flexible and more streamlined than ever before. Researchers can analyze all the data from an experiment at once, just importing the data files for all the participants of an eye-tracking experiment into one data viewer viewing session. You can set interest periods in Data Viewer to focus on critical temporal epochs in the data quickly and easily. These interest periods can be synced to stimulus onset markers from the experimental paradigm. Moreover, the interest period definition applies to all trials for all participants, making analysis quick and easy. We created 11 interest periods, where one of them represented the baseline and ten of them represented the ten words in the stimuli for each trial.

3.9. Pilot Study

First of all, a pilot study was conducted in order to detect and eliminate any problems with the eye-track stand and headphones in our laboratory. The pilot study was carried out at least three times to gain enough experience with the setup. The primary objective of the pilot study was to identify any potential issues within the experimental setup.

3.10. Data Preprocessing and Analysis

At the end of each trial, participants were asked to write three words that were spoken by the target speaker. If the words written were among those spoken by the target speaker, it was considered that the participant was displaying listening attention at the moment those words were being spoken. Overall accuracy was calculated based on the number of correct words, i.e., the matching words spoken by the target speaker. Also, some participants wrote words spoken by the competing speaker, rather than the target speaker (102 words in total). We have also isolated these moments of displaying attention to the competing speaker and included them in the analysis.

Pupil size and eye gaze data were analyzed for the one second before the speech onset and for the entire time the word lists were played. The eye-tracking data when participants were not listening, such as while entering the words with the keyboard, were filtered out. Since the starting and ending times of each short word was fixed and synchronized with the eye-tracking data, the recorded data could be associated with the words spoken at that moment. During the experiment, for each participant, the horizontal coordinates of eye positions (Fixation X), vertical coordinates of eye positions (Fixation Y), rapid horizontal movements (Saccade X), rapid vertical movements (Saccade Y), and the pupil dilation data were recorded. Horizontal and vertical coordinates of the participants' eye positions and pupil size changes in a certain period were obtained by using the EyeLink Data Viewer. The data was also analyzed using this program (SR Research Ltd., Mississauga, Canada).

For Fixation X, Fixation Y, Saccade X, and Saccade Y, we filtered the participants' data recorded precisely when they heard the words they had written, which ensured that they were listening during those specific moments. The rest of the data was not analyzed, since for other moments there was no way to tell whether the participants were displaying listening attention or not. Since 2 s were reserved for each word, mean values of Fixation X and Y, and Saccade X and Y recorded during those 2 s were calculated. The results were converted so that the origin is the center of the screen, rather than the top left of the screen. The first saccade data was taken when the first word was presented to each participant, whether or not they had written that word.

The value of the baseline pupil size in each trial was calculated separately for each participant by taking the mean of the participant's pupil size in the interval 1 s before the speech onset. The pupil size at the time each word was presented in each condition was calculated by taking the mean of the pupil size during 2 s duration for each word until the next word presentation. Then, the pupil dilation was calculated by subtracting the baseline pupil size from the pupil size for each word and condition for all participants one by one. Peak pupil size for each trial was defined as the participant’s

largest pupil size during the entire time speech was being played. The latency of peak pupil size was defined relative to the stimuli onset. It was converted to the order of the word being played at the moment the peak latency occurred.

Participant gender, age, handedness, target speaker gender, audio direction, and absence and presence of interference were the factors that were readily available. Among this data, instead of using participants' age as is, participants were divided into two age groups according to the median; young adults, between the ages of 19 and 25 and middle-aged adults, between the ages of 26 and 44.

3.11. Ethics

Research endeavors carried out at Middle East Technical University (METU) or undertakings led by METU personnel or students, involving the collection of data from human participants, are obligated to undergo scrutiny by the METU Human Subjects Ethics Committee (HSEC). We applied to the Human Subjects Ethics Committee to start data collection from human participants. According to METU HSEC policy, we will not store any of the private data of our participants for privacy concerns. The participants of our experiments were provided with the following information: the purpose of the study/experiment, the confidentiality of the data gathered from their involvement in the experiments, the anticipated duration of the experiments, and the location where the experiments would be conducted. The participants were also granted the freedom to withdraw from the experiment at their discretion. The Ethical Committee approval form can be found in Appendix B.

3.12. Assumptions of the Study

Within the context of this study, the following assumptions were taken into account:

- Participants provided accurate responses to the questionnaires and open-ended inquiries.
- The measurements employed in the study were deemed reliable and valid.
- The data recorded, collected, and analyzed were anticipated to be accurate.
- The study was expected to occur in a standard environment.

CHAPTER 4

EFFECTS OF SELECTIVE DIRECTIONAL AUDITORY ATTENTION ON EYE MOVEMENTS AND PUPIL RESPONSES

4.1. Research Questions of the Experiment

1. Does the level of listening attention have a relation with pupil size in processing speech?
2. Does interference have an effect on listening attention, and can this effect be determined through changes in pupil size?
3. Does listener gender affect auditory attention?
4. Does the source location of the sound affect listening to a task?

The purpose of this study was to measure the selective auditory attention given to the sound direction using eye movements and pupil responses considering gender differences. In this study, stimuli were presented to only the right ear, only the left ear, and both ears. The stimuli contained words spoken by either a male or a female speaker. We hypothesized that participants tend to look to the right when the target sound comes from the right and look to the left when it comes from the left. Therefore, horizontal and vertical level fixations and saccades of the eye gaze have been examined. Pupil dilations were recorded to measure auditory location attention, and baseline pupil size, peak pupil size, and pupil latency were calculated. Moreover, it was tested whether the participants' task performance decreased when there was location uncertainty in the presence of a competing talker. Thus, it is expected that the processing load will increase, the pupil dilation will be larger, and the performance of the participants will decrease.

4.2. Data Preprocessing and Analysis

Before the statistical analysis dependent and independent variables were defined. There were 10 dependent variables, which were Fixation X, Fixation Y, Saccade X, Saccade Y, Baseline Pupil Size, Pupil Size, Pupil Dilation (difference from baseline), Peak Pupil Size, Peak Latency, and Overall Accuracy (number of correct words). The independent variables were Audio Direction (left, right, bilateral), Participant Gender (male, female), Participant Age Group, Target Speaker Gender (male, female),

Interference (yes, i.e., multi-talker, no, i.e., single-talker), Focusing on Competing Talker (yes, no).

Firstly, taking into account the defined dependent and independent variables, an independent t-test, Mann Whitney test, Mann Whitley paired test and a repeated one-way ANOVA test were conducted and individual effects of the independent variables on the dependent variables were examined (Table A.1 in Appendix).

In this study, appropriate statistical tests were chosen based on the distribution and nature of the data. Specifically, a t-test was utilized when the data exhibited a normal distribution, allowing for the comparison of means between two groups. Conversely, when the assumption of normality was not met, the Mann-Whitney U test, a non-parametric test, was applied to compare distributions between two groups. Additionally, for paired samples where the data did not follow a normal distribution, the non-parametric paired counterpart, the Wilcoxon signed-rank test, specifically the Mann-Whitney U test for paired samples, was conducted. These statistical tests were selected to ensure robust and accurate analyses, considering the underlying characteristics and assumptions of the data, thereby enhancing the validity and reliability of the results obtained.

We ran linear mixed effect model analysis in R (4.2.1) using the lme4 package (Bates, Maechker, Bolker, & Walker, 2015). We included each dependent variable response as the dependent variable and added fixed effects of independent variables, as well as the interaction between these independent variables. We included Participant as a random effect. To estimate degrees of freedom and generate p values for mixed effect model, significance was calculated using the lmerTest package (Kuznetsova, Brockhoff, & Christensen, 2017) (Table A.2 in Appendix).

The model specification was as follows:

- Model: fixation_x ~ audio_direction * p_gender + (1 | subject)
- Model: fixation_y ~ interference * p_gender + (1 | subject)
- Model: saccade_x ~ 1 + audio_direction * p_gender + (1 | subject)
- Model: pupil_size ~ audio_direction * interference + p_gender + (1 | subject)
- Model: pupil_dilation ~ audio_direction * interference + p_gender + (1 | subject)
- Model: baseline_pupil ~ p_gender * age_group + (1 | subject)
- Model: peak_pupil ~ audio_direction * interference + p_gender + (1 | subject)
- Model: overall_accuracy ~ interference * s_gender + focusing_competing_speaker + (1 | subject)

4.3. Results

Analyses results are presented by grouping the dependent variables as eye movement related (Fixation X, Fixation Y, Saccade X, Saccade Y), pupil related (Baseline Pupil Size, Pupil Size, Pupil Dilation, Peak Pupil Size, Peak Latency) and Overall Accuracy. The results are presented for those with a p-value less than 0.05. All error bars indicate 95% confidence interval. Effect size (r) and eta square (η^2) results were presented.

Effect size results is crucial because it provides important information about the strength and direction of the relationship between variables. A higher absolute value of "r" indicates a stronger relationship, allowing for a quantitative assessment of the effect's magnitude. The presentation of eta squared (η^2) in research results serves the purpose of quantifying the effect size or the proportion of variance explained by an independent variable in a statistical analysis. Eta squared is a useful metric as it helps in understanding the practical significance or importance of the effects observed in a study. A larger eta squared indicates a more substantial effect size, implying that the independent variable has a greater impact on the dependent variable.

4.3.1. Eye Movement Related Results

The mean values of Fixation X, Fixation Y, Saccade X, and Saccade Y can be found in Table A.1, as well as the individual effects of independent variables on them.

4.3.1.1. Fixation Results

A repeated measures ANOVA with a Greenhouse-Geisser correction test was conducted to evaluate the effect of Audio Direction on Fixation X in left, bilateral and right directions. There was a significant effect of Audio Direction on Fixation X results, with F (1.38, 51.09) = 10.52, $p < .001$, $\eta 2 = 0.105$. Three post hoc paired-samples t-tests, corrected with Holm's sequential Bonferroni procedure, indicated that the mean score for the left direction (M = -70.49, SD = 131.35) was significantly different from that of the bilateral direction (M = -18.85, SD = 85.02) ;(t (37) = -2.725, $p < 0.05$), and that of the right direction (M = 26.28, SD = 128.8); (t (37) = -3.566, $p <$ .001). Also, the mean score for the bilateral direction was significantly different from that of the right direction with $p < 0.05$.

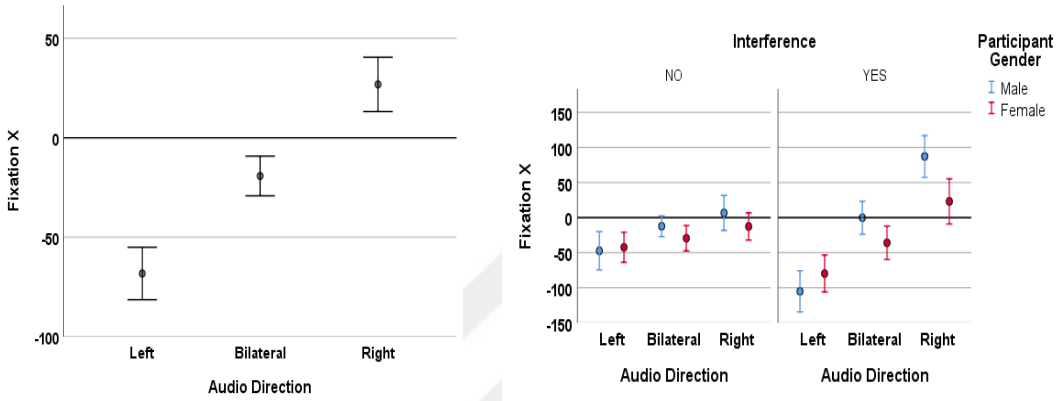
As a result of the Mann Whitney test which was conducted to evaluate Participant Gender differences in Fixation X, it was found that male participants' Fixation X (Mdn= -8.50) were significantly different from the female participants' (Mdn= -19.50); W = 1636656, $p < .001$, r = -0.03.

To examine the effect of Audio Direction and Participant Gender conditions on Fixation X, a logistic mixed-effect model was fitted on participants' pupil size, using Audio Direction and Participant Gender as fixed effect factors, with Subject (Participant) as random effect factors without data aggregation. Models were constructed using lme4 package in R. Fixed and random factors entered the model. They remained in the model if they significantly improved the model according to the Chi-square test ($p < 0.05$) and AIC (AIC is a single number score that can be used to determine which of multiple models is most likely to be the best model for a given dataset). The effects of each factor and their interactions were determined with the post-hoc Wald test.

There was a significant main effect of Audio Direction condition ($\chi 2 = 115.025$, df = 2, $p < .001$) and interaction between Audio Direction and Participant Gender conditions ($\chi 2 = 11.374$, $df = 2$, $p = 0.003$), but no main effect of Participant Gender ($\chi 2 = 0.263$, $df = 1$, $p = 0.608$). The Fixation X results were similar in males (M = -75.44, SD = 253.59) and females (M = -60.44, SD = 205.46) for the left Audio

Direction. However, for the right Audio Direction, Fixation X results were significantly greater in males (*M* = 47.22, *SD* = 243.89) than in females (*M* = 5.09, *SD* = 224.84). It was also the case for the bilateral Audio Direction. Fixation X results for bilateral Audio Direction were significantly greater in males (*M* = -6.65, *SD* = 166.02) than in females (*M* = -32.55, *SD* = 176.22).

The effects of Audio Direction, Interference, and Participant Gender on Fixation X are shown in Figure 12.



(a) **(b)**

Figure 12: (a) Effect of Audio Direction on horizontal eye positions (Fixation X) for left, bilateral and right audio directions. (b) Effect of Audio Direction, Interference and Participant Gender on Fixation X.

As a result of the Mann Whitney paired sample test which was conducted to evaluate Speaker Gender differences in Fixation X, it was found that in a male speaker condition, Fixation X (*Mdn*= -11.62) were significantly different from the female speaker condition (*Mdn*= -14.37*)*; *V = 598176*, *p* = 0.044, *r* = 0.052.

As a result of the Mann Whitney test which was conducted to evaluate Focusing on Competing Speaker effect in Fixation X, it was found that Fixation X in case of Focusing on Competing Speaker (*Mdn* = 4.45) was significantly larger than in the case of Focusing on Target Speaker (*Mdn* = -13.3); *W* = 145812, *p* = 0.032, *r* = 0.036.

As a result of the Mann Whitney test which was conducted to evaluate Age Group differences in Fixation X, it was found that Fixation X results in 19-25 Age participants (*Mdn* = -5.20) were significantly different from 26-44 age participants (*Mdn* = -20.57); *W* = 168152, *p* < .001, *r* = -0.10.

As a result of the Mann Whitney test which was conducted to evaluate Participant Gender differences in Fixation Y, it was found that male participants' Fixation Y (*Mdn*= 16.1) were significantly different from the female participants' (*Mdn*= 5.11*)* ;), *W* = 1603173, *p* < .001, *r* = -0.06.

As a result of the Mann Whitney test which was conducted to evaluate Age Group differences in Fixation Y, it was found that Fixation Y results for participants in Age

Group 19 to 25 (*Mdn* = 4.67) were significantly different those for the Age Group 26 to 44 (*Mdn* = 21.63), *W* = 1263328, *p* < .001, *r* = -0.14.

As a result of the Mann Whitney test which was conducted to evaluate Interference effect in Fixation Y, it was found that interference Yes Fixation Y (*Mdn*= 17.35) were significantly different from the interference No (*Mdn*= 5.39*); V* = 634976, *p* < .001, *r* = 0.09.

As a result of the Mann Whitney test which was conducted to evaluate Focusing on Competing Speaker effect in Fixation Y, it was found that Fixation Y in case of Focusing on Competing Speaker (*Mdn*= 12.6) was significantly larger than in the case of Focusing on Target Speaker (*Mdn*= -0.05); *W* = 187280, *p* = 0.038, *r= 0.035.*

4.3.1.2. Saccade Results

As a result of the Mann Whitney test which was conducted to evaluate Participant Gender differences in Saccade X, it was found that male participants' Saccade X (*Mdn*= -7.4) were significantly different from the female participants' (*Mdn*= -20.15*)*); *W* = 232021, *p* < .001, *r* = 0.118.

As a result of the Mann Whitney test which was conducted to evaluate Age Group differences in Saccade X, it was found that Saccade X results for participants in Age Group 19 to 25 (*Mdn* = -9.8) were significantly different from those for the participants in Age Group 26 to 44 (*Mdn* = -23.9); *W* = 222439, *p* = 0.006, *r* = 0.078.

A repeated measures ANOVA with a Greenhouse-Geisser correction test was conducted to evaluate the effect of Audio Direction on Saccade X in left, bilateral and right directions. There was a significant effect of Audio Direction on Fixation X results, with *F* (1.40, 51.93) = 8.29, *p* = 0.002, $\eta 2$ = 0.083. Three post hoc paired-samples t-tests, corrected with Holm's sequential Bonferroni procedure, indicated that the results for the left Audio Direction (*M* = -68.44, *SD* = 222.60) were significantly different from the right Audio Direction (*M* = -14.77, *SD* = 177.69) and bilateral Audio Direction (*M* = -21.07, *SD* = 174.61). However, the results for the right Audio Direction (*M* = -14.77, *SD* = 177.69) were similar to the bilateral Audio Direction (*M* = -21.07, *SD* = 174.61).

Figure 13 displays the effects of Audio Direction and Interference on Saccade X.

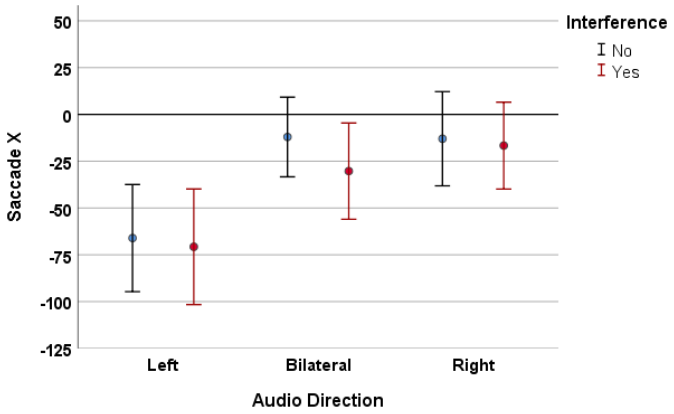


Figure 13: Saccade X (horizontal rapid eye movement) results recorded at the onset of each stimulus according to the direction of the target speech and the presence of interference.

4.3.2. Pupil Related Results

4.3.2.1. Baseline Pupil Size Results

Baseline Pupil Sizes of the participants were calculated as the average pupil diameter over one second preceding the onset of the stimuli.

As a result of the Mann Whitney test which was conducted to evaluate Participant Gender differences in Baseline Pupil Size. Baseline Pupil Sizes of female participants (*Mdn*= 4.77) were significantly larger that male participants (*M*dn= 4.25); W = 1073124, $p < .001$, r = 0.422. As a result of the Mann Whitney test which was conducted to evaluate Age Group differences in Baseline Pupil Size, it was found that Participants in the Age Group 19 to 25 (*Mdn* = 4.67) also had significantly larger Baseline Pupil Sizes than those in the Age Group 26 to 44 (*Mdn* = 4.34); W = 2621637, $p < .001$, r = 0.22.

Figure 14 shows Baseline Pupil Sizes in mm according to Age Group and Participant Gender.

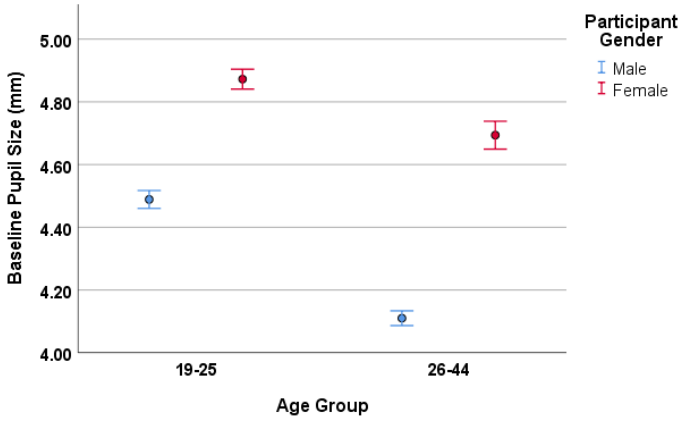


Figure 14: Baseline pupil sizes, calculated as the mean pupil size during the 1 s interval before the speech onset, according to Age Group and Participant Gender, showing that the Baseline Pupil Size decreases with age and female participants have larger baseline pupil sizes than male participants.

4.3.2.2. Pupil Size Results

As a result of the Mann Whitney test which was conducted to evaluate Participant Gender differences in Pupil Size, it was found that Female participants (*Mdn*= 5.07) had significantly larger Pupil Size than male participants (*Mdn*= 4.63); $W = 929882$, $p < .001$, $r = 0.327$. As a result of the Mann Whitney test which was conducted to evaluate Age Group differences in Pupil Size, it was found that the participants in the Age Group 19 to 25 displayed significantly larger Pupil Size (*Mdn*= 4.95) than those in the Age Group 26 to 44 (*Mdn= 4.63*); $W = 1863683$, $p < .001$, $r = -0.213$.

A repeated measures ANOVA with a Greenhouse-Geisser correction test was conducted to evaluate the effect of Audio Direction on Pupil Size in left, bilateral and right directions. There was significant effect of Audio Direction on Pupil Size results, with F (2, 74) = 7.153, $p = 0.001$, $\eta 2 = 0.003$. Three post hoc paired-samples t-tests, corrected with Holm's sequential Bonferroni procedure, indicated that the mean score for the bilateral direction ($M = 4.9$, $SD = 0.76$) was significantly different from that of the left direction ($M = 4.82$, $SD = 0.74$) (t (37) = -3.246, $p < 0.05$), and that of the right direction ($M = 4.83$, $SD = 0.73$) (t (37) = 2.832, $p < 0.05$). However, the mean score for the left direction was not significantly different from that of the right direction with $p = 0.451$.
As a result of the Mann Whitney test which was conducted to evaluate Interference effect in Pupil Size, it was found that Pupil Size in the case of Interference "Yes" *(Mdn* = 4.91*)* was significantly larger than in the case of Interference "No" *(Mdn* = 4.77*); V* = 377312, $p < .001$, $r = 0.295$.

To examine the effect of Audio Direction, Interference, and Participant Gender conditions on Pupil Size, a logistic mixed-effect model was fitted on participants' pupil size, using Audio Direction, Interference and Participant Gender as fixed effect factors, with Subject (Participant) as random effect factors without data aggregation. Models were constructed using lme4 package in R. Fixed and random factors were entered the model. They remained in the model if they significantly improved the model according to the Chi-square test ($p < 0.05$) and AIC (AIC is a single number score that can be used to determine which of multiple models is most likely to be the best model for a given dataset.). The effects of each factor and their interactions were determined with the post-hoc Wald test.

There was a significant main effect of Audio Direction condition ($\chi 2 = 11.191$, $df = 2$, $p = 0.004$), Interference condition ($\chi 2 = 63.769$, $df = 1$, $p < .001$), Participant Gender ($\chi 2 = 4.750$, $df = 1$, $p = 0.030$), and interaction between Audio Direction and Interference conditions ($\chi 2 = 23.703$, $df = 2$, $p < .001$). The Pupil Size results for the Interference "Yes" cases ($M = 4.7$, $SD = 0.54$) were significantly greater than for the Interference "No" cases ($M = 4.60$, $SD = 0.62$) in the left Audio Direction and Interference "Yes" cases ($M = 4.79$, $SD = 0.62$) were significantly greater than for the Interference "No" cases ($M = 4.56$, $SD = 0.59$) in the Bilateral Audio Direction for Male Participants. However, for the right Audio Direction, Pupil Size results for the Interference "Yes" cases ($M = 4.64$, $SD = 0.57$) were similar to Interference "No" cases ($M = 4.59$, $SD = 0.61$) for male participants. Otherwise, Male Participants' Pupil Size in the bilateral Audio Direction ($M = 4.79$, $SD = 0.62$) were significantly greater than left Audio Direction ($M = 4.70$, $SD = 0.54$) and Right Audio Direction ($M = 4.64$, *SD*

= 0.57) for the Interference "Yes" cases. The Pupil Size results for the Interference "Yes" cases (M = 5.15, SD = 0.87) were significantly greater than for the Interference "No" cases (M = 4.93, SD = 0.76) in the left Audio Direction for Female Participants. The Pupil Size results for the Interference "Yes" cases (M = 5.24, SD = 0.82) were significantly greater than for the Interference "No" cases (M = 5.06, SD = 0.80) in the Bilateral Audio Direction for Female Participants. However, for the right Audio Direction, Pupil Size results for the Interference "Yes" cases (M = 5.09, SD = 0.81) were similar to Interference "No" cases (M = 5.04, SD = 0.77) for female participants. Otherwise, Female Participants' Pupil Size in the bilateral Audio Direction (M = 5.24, SD = 0.82) were significantly greater than left Audio Direction (M = 5.15, SD = 0.87) and Right Audio Direction (M = 5.09, SD = 0.81) for the Interference "Yes" cases. Also, under the same Audio Direction and Interference condition, Participant Gender does not significantly affect Pupil Size. Male and Female participants pupil sizes were similar.

Figure 15 shows Pupil Size results according to the Audio Directions, Interference, Participant Gender and Age Groups.

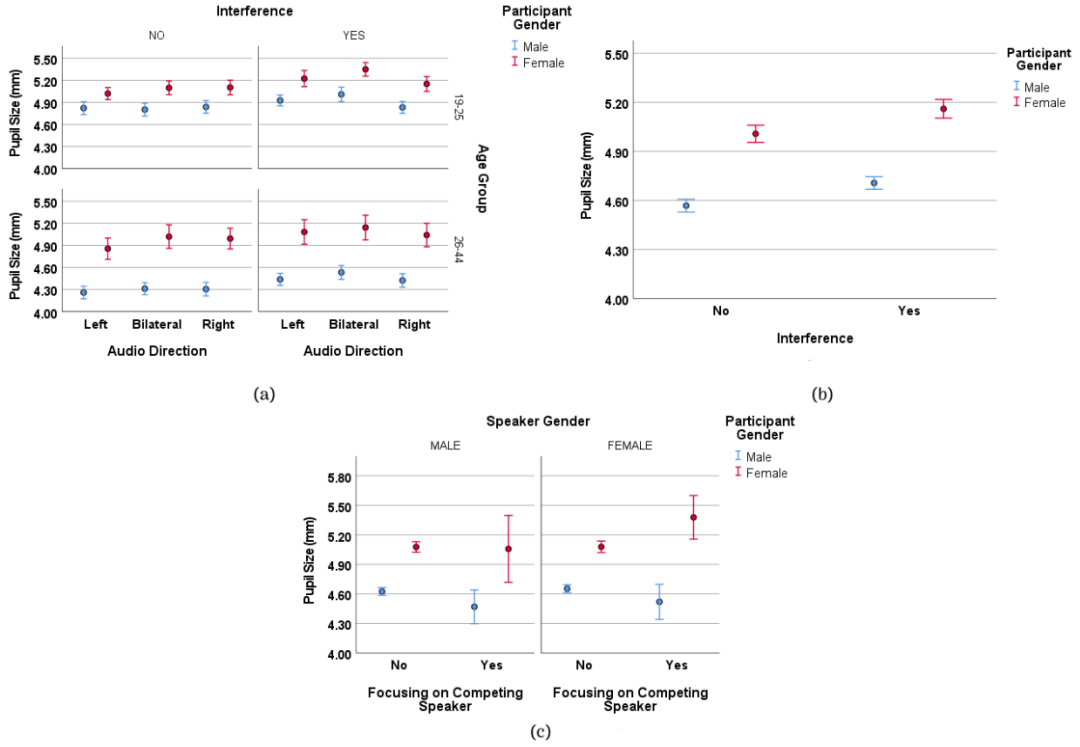


Figure 15: Pupil Size results calculated as the mean pupil sizes during the 2 s presentation of the word listened to by each participant. (a) According to Audio Direction and presence of Interference for different Age Groups, showing that there was a significant difference in Pupil Size results according to Audio Direction and presence of Interference, and that female participants had larger Pupil Sizes than male participants, and that as the age of the male participants increased, their pupil size decreased. (b) According to presence of Interference for Participant Gender, showing that the presence of Interference causes a significant increase in the Pupil Size results (c) According to whether the participants Focused on the Competing Speaker or not and the Speaker Gender for Participant Gender, showing that the Pupil Size results of the participants who focused on the competing speaker instead of the target speaker differed according to the gender of the participant.

4.3.2.3. Pupil Dilation Results

As a result of the Mann Whitney test which was conducted to evaluate Participant Gender differences in Pupil Dilation, it was found that Male participants (*Mdn* = 0.31) had significantly larger Pupil Dilation than female participants (*M*dn = 0.27); W = 1594001, $p < .001$, r = 0.06.

As a result of the Mann Whitney test which was conducted to evaluate Age Group differences in Pupil Dilation, it was found that the participants in the Age Group 19 to 25 displayed significantly larger Pupil Dilation (*Mdn* = 0.32) than those in the Age Group 26 to 44 (*Mdn* = 0.27); W = 1601029, $p < .001$, r = -0.06.

A repeated measures ANOVA with a Greenhouse-Geisser correction test was conducted to evaluate the effect of Audio Direction on Pupil Dilation in left, bilateral and right directions. There was not significant effect of Audio Direction on Pupil Dilation results, with F (1.60, 59.27) = 2.78, p = 0.081, $\eta 2$ = 0.008.

As a result of the Mann Whitney test which was conducted to evaluate Interference effect in Pupil Dilation, it was found that Pupil Dilation in the case of Interference "Yes" *(Mdn* = 0.37*)* was significantly larger than in the case of Interference "No" *(Mdn* = 0.22*);* V = 348753, $p < .001$, r = 0.332.

As a result of the Mann Whitney test which was conducted to evaluate Focusing on Competing Speaker effect in Pupil Dilation, it was found that Pupil Dilation in the case of Focusing on Competing Speaker (*Mdn* = 0.37) was significantly larger than in the case of Focusing on Target Speaker (*Mdn* = 0.29); W = 136987, p = 0.007, r = 0.046.

Figure 16 shows that the Pupil Dilation results in mm, according to the Audio Directions, Interference, Participant Gender and Age Groups.

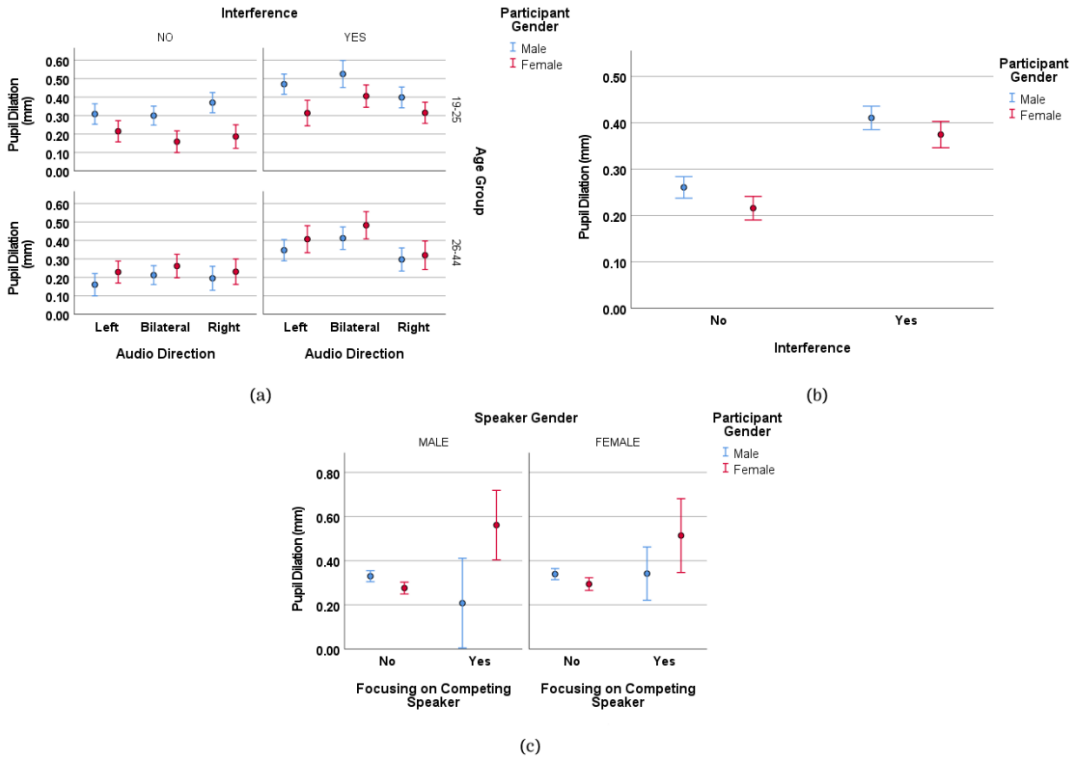


Figure 16: Pupil Dilation results, calculated by subtracting the baseline pupil size from the pupil size for each word and condition for all participants individually. (a) According to Audio Direction and presence of Interference for different Age Groups, showing that there was a significant effect of Audio Direction on the Pupil Dilation, Pupil Dilation results of male participants in the Age Group 19 to 25 were significantly greater than for the male participants in the Age Group 26 to 44, and the effect of Age Group on Pupil Dilation for female participants was not significant. (b) According to presence of Interference for Participant Gender, showing that in cases of Interference, both male and female participants' Pupil Dilation increased, and that the male participants' Pupil Dilation was larger than female participants'. (c) According to whether the participants Focused on the Competing Speaker or not and the Speaker Gender for Participant Gender showing that there is a significant difference between the Pupil Dilation results of female participants who Focused on Competing Speaker instead of Focusing on Target Speaker regardless of the speaker gender.

To examine the effect of Audio Direction, Interference, and Participant Gender conditions on Pupil Dilation, a logistic mixed-effect model was fitted on participants' pupil dilation, using Audio Direction, Interference and Participant Gender as fixed effect factors, with Subject (Participant) as random effect factors without data aggregation. Models were constructed using lme4 package in R. Fixed and random factors were entered the model. They remained in the model if they significantly improved the model according to the Chi-square test ($p < 0.05$) and AIC (AIC is a single number score that can be used to determine which of multiple models is most likely to be the best model for a given dataset). The effects of each factor and their interactions were determined with the post-hoc Wald test.

There was a significant main effect of Interference condition ($\chi2 = 69.554$, $df = 1$, $p < .001$), and interaction between Audio Direction and Interference conditions ($\chi2 = 27.367$, $df = 2$, $p < .001$). However, there was not significant main effect of Audio Direction condition ($\chi2 = 0.993$, $df = 2$, $p = 0.609$), Participant Gender ($\chi2 = 0.435$, $df = 1$, $p = 0.510$). The Pupil Dilation results for the Interference "Yes" cases ($M = 0.41$,

SD = 0.35) were significantly greater than for the Interference "No" cases (*M* = 0.24, *SD* = 0.37) in the left Audio Direction, Interference "Yes" cases (*M* = 0.47, *SD* = 0.41) were significantly greater than for the Interference "No" cases (*M* = 0.26, *SD* = 0.33) in the Bilateral Audio Direction for Male Participants and for the right Audio Direction, Pupil Dilation results for the Interference "Yes" cases (*M* = 0.35, *SD* = 0.37) were significantly greater than Interference "No" cases (*M* = 0.29, *SD* = 0.38) for male participants. Otherwise, Male Participants' Pupil Dilation in the bilateral Audio Direction (*M* = 0.47, *SD* = 0.41) were significantly greater than left Audio Direction (*M* = 0.41, *SD* = 0.35) and Right Audio Direction (*M* = 0.35, *SD* = 0.37) for the Interference "Yes" cases. The Pupil Dilation results for the Interference "Yes" cases (*M* = 0.36, *SD* = 0.43) were significantly greater than for the Interference "No" cases (*M* = 0.22, *SD* = 0.36) in the left Audio Direction for Female Participants. The Pupil Dilation results for the Interference "Yes" cases (*M* = 0.45, *SD* = 0.40) were significantly greater than for the Interference "No" cases (*M* = 0.21, *SD* = 0.37) in the Bilateral Audio Direction for Female Participants. Moreover, for the right Audio Direction, Pupil Dilation results for the Interference "Yes" cases (*M* = 0.32, *SD* = 0.41) were similar to Interference "No" cases (*M* = 0.21, *SD* = 0.41) for female participants. Otherwise, Female Participants' Pupil Size in the bilateral Audio Direction (*M* = 0.45, *SD* = 0.40) were significantly greater than left Audio Direction (*M* = 0.36, *SD* = 0.43) and Right Audio Direction (*M* = 0.32, *SD* = 0.41) for the Interference "Yes" cases. Also, under the same Audio Direction and Interference condition, Participant Gender does not significantly affect Pupil Dilation. Male and Female participants pupil dilation were similar.

4.3.2.4. Peak Pupil Size Results

As a result of the Mann Whitney test which was conducted to evaluate Participant Gender differences in Peak Pupil Size, it was found that Female participants (*Mdn*= 5.64) had significantly larger Pupil Size than male participants (*Mdn* = 5.2); *W* = 937351, $p < .001$, *r* = 0.323.

As a result of the Mann Whitney test which was conducted to evaluate Age Group differences in Peak Pupil Size, it was found that the participants in the Age Group 19 to 25 displayed significantly larger Pupil Size (*Mdn* = 5.55) than those in the Age Group 26 to 44 (*Mdn*= 5.23); *W* = 1863754, $p < .001$, *r* = -0.213.

A repeated measures ANOVA test was conducted to evaluate the effect of Audio Direction on Peak Dilation in left, bilateral and right directions. There was a significant effect of Audio Direction on Peak Pupil Size results, with $F(2, 74) = 6.059$, $p = 0.004$, $\eta 2 = 0.002$. Three post hoc paired-samples t-tests, corrected with Holm's sequential Bonferroni procedure, indicated that the mean score for the bilateral direction (*M* = 5.46, *SD* = 0.79) was significantly different from that of the left direction (*M* = 5.39, *SD* = 0.78) ($t(37) = -3.344$, $p < 0.05$). However, the mean score for the right direction (*M* = 5.4, *SD* = 0.78) was not significantly different from that of the left direction and with bilateral direction $p > 0.05$.

As a result of the Mann Whitney test which was conducted to evaluate Interference effect in Peak Pupil Size, it was found that Peak Pupil Size in the case of Interference "Yes" *(Mdn* = 5.47*)* was significantly larger than in the case of Interference "No" *(Mdn* = 5.35*); V* = 700316, $p < .001$, *r* = 0.288.

Figure 17 shows the Peak Pupil Size results according to Audio Direction, Interference and Participant Gender.

Peak latency converted to the order of the word listened to in each trial was different for male and female participants as shown in Figure 18.

Female participants have larger Peak Pupil Size values and the Peak Pupil Size was reached while listening to the 7th and 8th words.

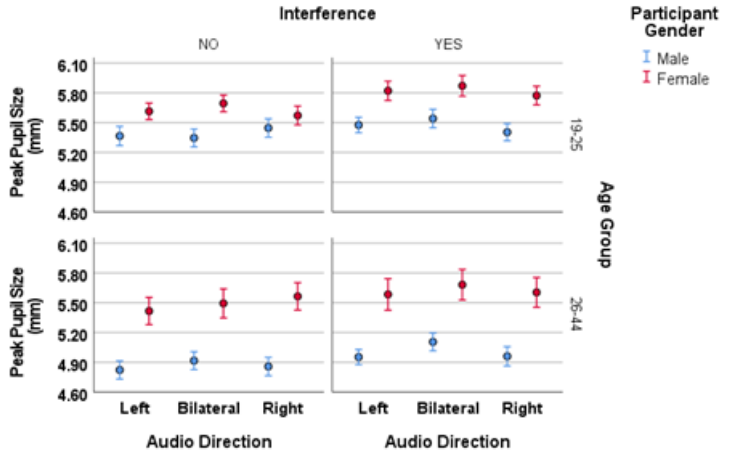
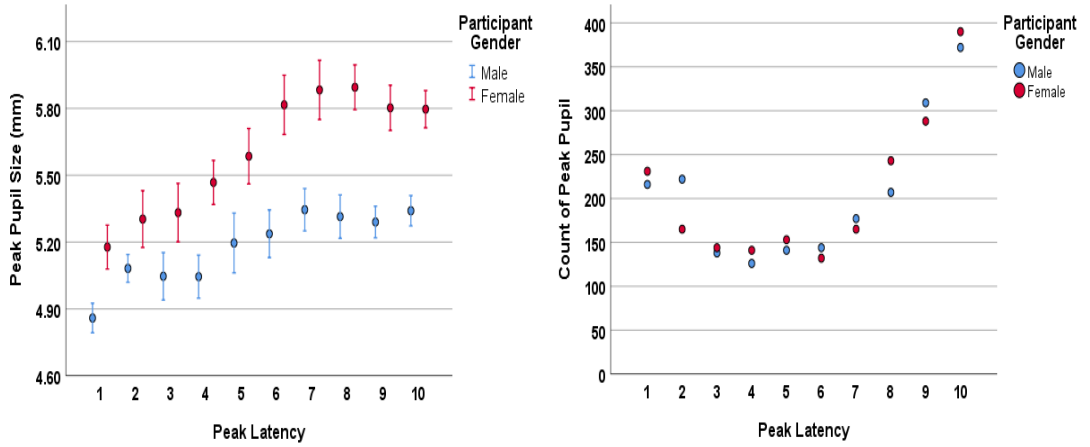


Figure 17: Peak Pupil Size, defined as the participants' largest pupil size during the entire time each trial was being listened to, according to the Audio Direction and Interference for Participant Gender, showing the Peak Pupil Size was greater in the presence of Interference, and Peak Pupil Size results of female participants were significantly larger than the male participants, and the participants in the Age Group 19 to 25 were found to have Larger Peak Pupil Size results than the participants in the Age Group 26 to 44.



(a) **(b)**

Figure 18: (a) Peak Pupil Size in mm according to the Peak Latency and Participant Gender showing that the female participants have larger Peak Pupil Size values and the Peak Pupil Sizes reached while listening to the 7th and 8th words were the largest (b) Count of the Peak Pupil Size cases according to the Peak Latency and Participant Gender, showing that the Peak Pupil Size was reached mostly while listening to the last words of each trial.

4.3.3. Overall Accuracy Results

Each of the 38 participants was expected to enter maximum three words for each of the 36 trials. This corresponds to 4104 expected words in total. Out of the 3579 words were written by the participants, 3470 words were correct, i.e., actually spoken by the target speaker. Accordingly, the mean overall accuracy, defined as the number of correctly written words, was *2.54 with SD* = 0.71.

Spearman's rank correlation was computed to assess the relationship between overall accuracy and pupil dilation. There was a significant positive correlation between the two variables, *(r_s* (36)) = 0.37, *95% CI [0.05, 0.63], p* = 0.010, *n*= 38. The result of the Pearson correlation indicated that there was a significant positive association between Pupil Dilation and the Overall Accuracy *t* = 3.75, *p* < .001, Pearson's (*r* (17)) = 0.67, *n* = 19, *95% CI [0.38, 1.00]* of male participants. However, the result of Spearman's rank correlation indicated that there was no statistically significant positive association between the Pupil Dilation and the Overall Accuracy *(r_s* (17)) = 0.12, *95% CI [-0.36, 0.56], p* = 0.306, *n* = 19 of female participants.

As a result of the two-sample t test which was conducted to evaluate Participant Gender differences in Overall Accuracy, it was found that Male participants (*M*= 2.69, *SD*= 0.69*)*) were significantly more successful than female participants (*M* = 2.48, *SD* = 0.76); *t (*4102) = 4.5612, *p* < .001.

A two-sample t-test which was conducted to evaluate Age Group differences in Pupil Size, it was found that the participants in the Age Group 26 to 44 were significantly more successful (*M* = 2.57, *SD* = 0.74) than those in the Age Group 19 to 25 (*M* = 2.51, *SD* = 0.72); *t (*4102) = -2.63, *p* = 0.009.

A repeated measures ANOVA test was conducted to evaluate the effect of Audio Direction on Overall Accuracy in left, bilateral and right directions. There was a significant effect of Audio Direction on Overall Accuracy results, with *F* (2, 74) = 3.481, *p* = .036, $\eta 2$ = 0.032. Overall Accuracy for the left Audio Direction (*M* = 2.60, *SD* = 0.72) was significantly greater than the bilateral Audio Direction (*M* = 2.48, *SD* = 0.76) and the right Audio Direction (*M* = 2.53, *SD* = 0.71). However, the Overall Accuracy results were similar for bilateral and right Audio Directions.

As a result of the Mann Whitney test which was conducted to evaluate Interference effect in Overall Accuracy. The Overall Accuracy in the case of Interference "Yes" (*Mdn* = 2.44) were smaller than in the case of Interference "No" (*Mdn* = 2.72); *V* = 641.5, *p* < .001, *r* = 0.715.

A paired t test was conducted to evaluate Speaker Gender effect in Overall Accuracy. It was found that the participants were significantly more successful while listening to the male speaker (*M*= 2.59, *SD*= 0.7) than listening to the female speaker (*M*= *2*.48, *SD*= 0.75); *t (*2051) = 4.824, *p* < .001.

A two-sample t-test which was conducted to evaluate Focusing on Competing Speaker effect in Overall Accuracy. It was found that participants who focus on Target Speaker

(M= 2.584, SD= 0.66) was more successful than those who focus on the Competing Speaker (M= 0.64, SD= 0.77); t (4102) = 29.202 p < .001.

Figure 19 shows how the Overall Accuracy changes according to the Audio Direction, presence of Interference and Participant Gender, as well as for the Speaker Gender. Figure 20 shows the frequency of the order of the words in the stimuli that were written.

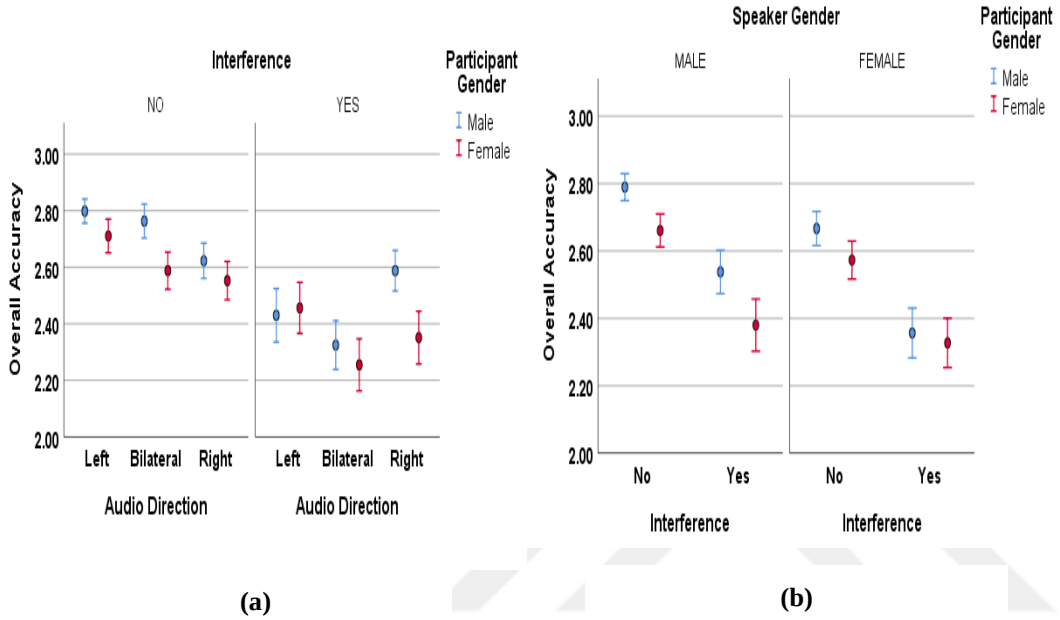


Figure 19: Overall Accuracy of the participants taken as the number of correct words, i.e., spoken by the target speaker (a) based on the Audio Direction, presence of Interference and Participant Gender, showing that in the absence of interference, participants were more successful in listening situations where the sound was given from the left than from the right. (b) based on the presence of Interference, Speaker Gender and Participant Gender, showing that the male participants were more successful than female participants, and this difference was more when the speaker was male.

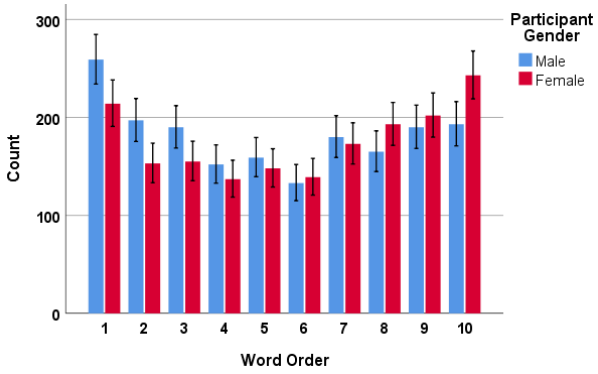


Figure 20: Frequency of the order of the words in the stimuli that were written, showing that the first 3 and the last 3 words of the stimuli were entered more than the other words of the stimuli by the participants.

4.4. Discussion

The purpose of the study was to investigate directional auditory attention and measure how this attention affects the eye movements and the pupil size with and without interfering speech. Moreover, this study investigated whether the gender of the speaker, gender of the listener and age affect this directional auditory attention.

To summarize the main findings, it can be said that when the listening attention is towards the sound's direction, an eye movement occurs towards that direction. Female and male participants' auditory attentions differ from each other, and this difference is reflected both in the pupil dilation and the horizontal eye movements. Since there were only two speakers, the effect of the speaker gender on the auditory attention could not be evaluated. Finally, interfering speech affects the directional auditory attention and this effect is different for female and male participants.

The following sections discusses the main findings.

4.4.1. Audio Direction Effects

According to Gopher (1973), the eye movement during selective auditory attention displays a pattern shifting towards the ear exposed to the sound and large saccades and long fixations occur towards the direction of the relevant ear. Our study also showed that the participants make eye movements towards the direction of the sound. In other words, when the sound comes from the left direction, the eyes move leftward, and when it comes from the right direction, the eyes move rightward. These eye movements also show similar results to the previous study of Hendrikse et al. (2018), which indicated that the participants quickly make eye gaze towards the speakers' direction and the study of Vertegaal et al. (2001) which indicated that 88% of the participants made eye gaze towards the speaker they listened to. In our study, although the participants made horizontal eye movements towards the sound source's direction, vertical eye movements relative to the audio direction were not found significant (Table A.1). Our study also showed a significant bias towards the left direction, which means the shift towards the left was more than the shift towards the right direction as observed for both the fixations and the saccades.

Significant statistical differences were found in the participants' pupil sizes concerning the direction of the sound. To illustrate, when participants listened to the sounds coming from both sides, the participants had larger pupil sizes. Also, during bilateral listening, the participants' pupil dilation had a significant increase in the presence of interference and their peak pupil sizes were larger.

In the past, many studies reported right ear advantages in dichotic listening. Hendrikse et al., (2018) stated that doing an eye movement to the right affected the right ear advantage. Moreover, Grange et al. (2016) stated that presenting stimuli to the right ear opposite to the left ear improves recall. This study's results are not in agreement with Hendrikse et al. (2018) and Grange et al. (2016). The participants' responses showed that presenting the stimuli to the left ear without interfering speech improves the task performance. In the absence of any direction, it has been observed that there

is a right ear advantage in dichotic listening (Westerhausen & Kompus, 2018). However, the advantage of the right ear can change when the listeners are asked to pay attention to one ear. As Asbjornsen & Hugdahl (1995), Foundas et al. (2006), Kompus et al. (2012), and Westerhausen & Kompus (2018) have already mentioned, it is seen that if the listener has attention to the right side, there is a right ear advantage, and if there is attention to the left side, there is a left ear advantage.

According to the results, the horizontal and vertical fixations reveal where the selective attention is directed to. In the presence of competing speech, the differences in horizontal fixations are more pronounced and can even be used to predict which speaker a participant is listening to. However, it should be noted that in this study, the speeches were presented to the left and right ears alone. If binaural listening conditions were to be tested, both ears would receive both speeches at varying interaural time and amplitude differences according to the directions of the sound sources. In such cases, differences in the horizontal fixations may not be as pronounced.

In our study, participants were informed which speaker was the target speaker that they should be listening to by mentioning the gender of the speaker without any cue on its direction. Kidd et al. (2005) stated that the cue about the target speaker had a positive effect on the task performance results. However, the effect of the cue about the target speaker in our study cannot be determined, since we did not have any trial without this cue. If the instruction had been the direction of the target speaker, rather than the gender, both the task performance and the eye movements could differ.

In our study, participants were instructed to attend to only the target speaker when there is competing speech, i.e., utilizing selective attention. If however, participants had been instructed to attend to both speakers, and write down the words spoken by any speaker, the audio direction could have different affects.

4.4.2. Participant Gender Effects

According to the results of the horizontal eye movements, the participants' first saccades recorded at the stimuli onset, varied according to the gender of the participant. Male participants looked towards the right more than the female participants for all three sound directions. Also, based on the participants' fixations on the horizontal plane during the listening, it can be said that the participant gender affects the fixations made towards the direction of the sound. For the cases where the sound played bilaterally and, on the right, male participants fixated towards the right more than the female participants.

Baseline pupil sizes measured before starting the trials also varied according to the gender of the participants. It was observed that the female participants' baseline pupil sizes were larger than the male participants' baseline pupil sizes. Moreover, pupil sizes measured at the time of listening, as well as the peak pupil sizes were larger for female participants than male participants. Interestingly however, male participants had larger pupil dilations than female participants. Gilzenrat et al. (2010) interpreted this situation as a sign of better performance detected during auditory oddball tasks. Similarly, in our study, the performance of male participants was better than female participants.

There are different opinions about the effect of the listener gender on the pupil dilation in the literature. For example, Partala and Surakka (2003) stated that women had larger pupil dilations than men only for neutral sounds. On the contrary, Gingras et al., 2015) indicated that males had larger pupil dilations to music excerpts. Moreover, while the gender did not affect the pupil dilation in the study of Burley et al. (2017), the relationship between testosterone level and pupil dilation was mentioned by Dabbs (1997).

4.4.3. Interference Effects

The statistical comparison of horizontal eye positions in the presence of competing speaker showed that the participants' eye movements shifted leftward regardless of the sound's direction. In addition, female participants made more leftward eye movements than male participants. Furthermore, there was a significant difference in vertical eye movements with and without competing speaker. Upward eye movements were observed in the case of interference.

The pupil sizes, pupil dilations and the peak pupil sizes were larger in the presence of competing speaker. These results are in line with previous studies (Koelewijn et al., 2012, 2014; Zekveld et al., 2014, 2018). In Zekveld et al. (2018), this situation is interpreted as a sign of greater cognitive processing load caused by the distractor sentence.

Larger pupil dilations also occurred when participants displayed attention to the competing speaker, rather than the target speaker. Koelewijn et al. (2014) reported that in situations where two sentences were presented instead of one, and participants were required to report both sentences as opposed to just one, an increase in pupil size was observed.

When there is interference, presenting speeches bilaterally results in larger pupil dilation than presenting target and competing speeches to different ears. Since an increase in pupil dilation indicates more effort, it can be said that selective attention becomes easier when directional cues are present.

Hyönä et al. (1995) stated that the peak latency was the longest for the dual-target condition. However, in our study, there was only one target and it was found that the peak latencies in the case of interference were shorter than in the case of no interference; yet this result was not statistically significant.

4.4.4. Age Effects

Although our results are statistically meaningful, it seems not easy to conclude the effects of age on auditory location attention, since our experiment was not designed based on age differences. The age range of participants was between 19 and 44 years. Participants in the age group 19 to 25 looked towards the right more than the participants in the age group 26 to 44. Also, the former looked up more than the latter.

We found that as the age increased, the baseline pupil sizes shrunk. This result coincides with the results of previous studies covering the age range from 25 to 75

years (Morris et al., 1997; Zekveld et al., 2011). Young adults (between the ages 19 and 25) had larger pupil sizes than middle age adults (between the ages 26 and 44). Our results showed that different age groups had different pupil dilations during listening. Younger group had larger pupil dilations, which, as Piquado et al. (2010) mentioned, indicates a relationship between the age and the small dynamic range of the pupils. Also, Koch and Janse (2016) stated that the pupil dilation varies with age; older adults had smaller pupil dilation than younger adults. Moreover, younger participants had larger peak pupil sizes than older participants. On the other hand, the latter had better task performance than the former.

Our results are not consistent with the studies of Ayasse et al. (2017); Chaney et al. (1989); Koelewijn et al. (2012); and Kuchinsky et al. (2016), which found no relationship between the age and the pupillary response. Zekveld et al. (2018) noted that the reason for having two different opinions in the literature might be due to the narrow ranges of ages in these studies. They suggested that the researchers should test older participants.

4.5. Chapter Conclusion

This part of the study investigated the effects of selective directional attention on the eye movements and pupil responses. A test was designed that enabled filtering the captured eye movements and pupil responses at the exact moments of displaying selective directional attention. It was found that the subjects are more attentive to the stimuli presented to a single ear than to both ears, and the task performance increases accordingly. Though both females and males look towards the direction of the sound source, the extent of it differs according to the age and gender of the listener. The presence or absence of interference in multi-talker speech has a significant effect on the pupil dilation, as well as the age, gender and audio presentation direction. There is an inverse relationship between the pupil dilation and the overall accuracy.

One of the limitations of this study was that there were only two speakers. Despite the fact that they were both certified narrators, their pronunciations could still have affected the results. Another limitation was that the audio directions were only left, right and bilateral. This design might have magnified the effect of directional attention, which might be limited in normal binaural listening. There were only two simultaneously active speakers and more than two speakers were not considered.

Speech communication in daily life is quite complex and dynamic, because of the presence of multiple sound sources. In a cocktail party, listening is more difficult for listeners with hearing loss than for normal hearing listeners. As Litovsky et al., (2017) stated, if the listeners can separate the target speaker’s speech from interference, they can grasp more of the content of the speech. In this study, it was found that the listeners made eye movements towards the direction of the sound and pupil dilations were larger in the presence of a competing speaker. As Zekveld et al. (2018) and Van Engen & McLaughlin (2018) mentioned, the application areas of eye movement and pupillometry are increasing and the results of this study can contribute to these areas. The use of eye movement and pupillometry to capture the user's selective auditory attention in combination with different measures can also help us understand the needs

of listeners (Tessendorf et al., 2011; Zhang et al., 2021). As suggested by Alhanbali et al. (2021), pupil diameter can be used for better hearing rehabilitation with hearing impairment devices. The attentive hearing aid proposed by Tessendorf et al. (2011) is an example of such an application, which utilizes a wearable eye tracker with a hearing aid. When the hearing-impaired user looks at one of the sources, the user can only hear the sounds coming from that source by filtering the others. Thus, hearing impaired users can chat in a cafe with music in the background, or focus on their teacher in a noisy classroom. Moreover, as Ledoux et al. (2016) mentioned, it can be found out whether a word was understood or not by using pupil diameter of a listener. In this way, we can observe the language development of individuals with autism spectrum disorder and communication deficiencies, and increase their communication skills.

This study presented some of the factors and their interactions that need be considered for the ultimate aim of predicting what a person might be paying attention to by measuring the eye movements and the pupil responses. Future work will include designing new tests to extend this study with more speakers, binaural listening conditions and multiple sound sources. Further studies should investigate directional selective auditory attention using an eye tracker integrated into VR headsets.

CHAPTER 5

UNRAVELING AUDITORY ATTENTION AND PUPIL RESPONSES: INVESTIGATING DIFFERENCES BETWEEN HEARING AND LISTENING, SEMANTIC IMPACT, AND STIMULUS FAMILIARITY

In this study, we investigate auditory attention through eye movements and pupil dilation, aiming to discern the difference between hearing and active listening. We explore the impact of Speech Intelligibility Index (SII), Perceptual Evaluation of Speech Quality (PESQ), and Signal-to-Noise Ratio (SNR) on listening attention and pupil responses. Additionally, we examine how semantic content influences auditory attention, as reflected in pupil dilation, and how stimulus familiarity shapes these responses. Through these inquiries, we aim to deepen our understanding of auditory cognition and its interplay with physiological markers. In pursuit of this objective, novel methodologies were employed to address the outlined research inquiries.

5.1. Research Questions of the Experiment

1. Can we investigate the difference between hearing and listening using eye movements?
2. What is the relationship between the effect of SII (Speech Intelligibility Index), PESQ (Perceptual Evaluation of Speech Quality), and SNR (Signal-to-Noise Ratio) values on listening attention and pupil responses?
3. Does the semantic content have an impact on auditory attention, and how does this impact manifest in pupil dilation?
4. How does the familiarity of stimuli presented within the study influence auditory attention and pupil dilation?

5.2. Methods

5.2.1. Word Frequencies in Turkish

Word frequency is a measure of how often a word is used in a language (Balota and Chumbley, 1984; Gernsbacher, 1984; Monsell et al., 1989, Kuchinke, 2007). It been

shown to affect the speed and accuracy of word recognition, eye fixations during reading (Rayner and Duffy, 1986), and event-related potentials (Dambacher et al., 2006; Hauk and Pulvermüller, 2004; Rugg, 1990; Sereno et al., 1998). High frequency words are easier to recognize and process than low frequency words, which require additional analysis such as phonological processing (Coltheart et al., 2001). Word frequency is also a good predictor of word difficulty, as difficult words appear less often and words that are more commonly encountered are learned faster and recalled better (Breland, 1996). Word familiarity is also related to difficulty, as the more familiar a word is to an individual, the easier it may be for them to process (Dupuy, 1974; Chapman, 2015).

To find the frequency of use of each word included in the subjective tests conducted in this study Turkish National Corpus was used (Aksan et. al. 2012). Turkish National Corpus (TNC) is designed to be a balanced, large scale (50 million words) and general-purpose corpus for contemporary Turkish. It has benefited from previous practices and efforts for the construction of corpora. In this sense, TNC generally follows the framework of British National Corpus, yet necessary adjustments in corpus design of TNC are made whenever needed. All throughout the process, different types of open-source software are used for specific tasks, and the resulting corpus is a free resource for non-commercial use. TNC with a size of 50 million words, is a balanced and a representative corpus of contemporary Turkish. It consists of samples of textual data across a wide variety of genres covering a period of 24 years (1990-2013).

5.2.2. Emotional Sentiment Analysis

In a study conducted by Challis and Krane in 1988, it was found that both positive and negative words were processed more quickly. However, in a later study by Kuchinke et al. in 2005, it was found that only positive words were processed more quickly, and this result was supported by a meta-analysis of seven lexical decision task (LDT) studies by Siegle et al. in (2002) that found no difference in the processing of negative and neutral words. In a study by Siegle et al. (2001), the pupillary responses of depressed and non-depressed subjects were examined while they completed a lexical decision task using emotionally charged words. No effects of the emotional valence of the words were reported for the non-depressed subjects, but the depressed and non-depressed subjects differed on two of five identified factors. The study did not report on measures of peak pupil dilation that were evoked by the task. Kuchinke et al.'s 2005 study, which used functional magnetic resonance imaging (fMRI), also found that positive words led to faster reaction times and fewer errors compared to neutral and negative words in both a behavioral pilot study and the subsequent fMRI study. Furthermore, the brain regions associated with the processing of positive and negative words were found to be distinct, with the anterior prefrontal cortex being specifically associated with the processing of positive words. Moreover, Kuchinke et al. (2007) proposed that emotionally charged words do not lead to greater pupil dilations compared to neutral words, regardless of the frequency of the words.

The sentiment analysis of Turkish words (which was used in the study) was calculated using MATLAB. SWNetTR.csv (containing a total of 37K Turkish words) (Sağlam, 2019) was used to determine the tone and polarity of each word in our word list (saved in an excel file "word_with_tone_and_polarity.xlsx"). The polarity is designated as

either +1, signifying a positive sentiment, or -1, denoting a negative sentiment. Conversely, the tone has the capacity to exhibit continuous variations across the spectrum encompassing these two extremes. As such, the tone yields additional insights into the intensity of the sentiment, whereas the polarity delivers a concise and distinct overview.

5.2.3. Word Embedding (Word2Vec) Analysis

Word embeddings is a way of representing words as numerical vectors, which can capture various characteristics of the word in relation to the text it appears in (Bengio, 2003). This representation can be learned in a way similar to training a neural network. With these numerical representations, you can compare the similarity or dissimilarity between words. Word2Vec is a specific method for generating word embeddings, which is effective because it groups together vectors of similar words (Mikolov, 2013). Given a large enough dataset, Word2Vec can make strong estimates about a word's meaning based on its occurrences in the text, and can generate associations between words. You can also perform algebraic operations on word embeddings to find approximate similarities between words. For example, subtracting the embedding for "man" from the embedding for "king" and adding the embedding for "woman" gives a result that is close to the embedding for "queen".

Turkish Word to Vector analysis (Word2Vec) was conducted for the word list used in this study by Python. The WikiCormus module from the gensim.corpora library was used to fetch vectors of our word list from "trwiki-20220301-pages-articles-multistream.xml.bz2" downloaded from FastText (Facebook Open Source) (Word vectors for 157 languages, 2023). This webpage provide word vectors was trained on Common Crawl and Wikipedia using FastText. The training utilized CBOW with position-based weighting, employing 300 dimensions, character n-grams of length 5, a window of 5, and 10 negative instances. Using this trained word vectors for Turkish, a excel file was containing each word's vectors was created to be used later in MATLAB. Each word has 300 different vector values. The distances between the words were calculated. Agglomerative hierarchical cluster tree was used to compute the distance between clusters. To visualize the hierarchical cluster tree, the Dendrogram function of MATLAB was applied. As a result, it was decided to divide the word list into four categories as shown in Table 3.

Table 3: Vector group of Speech Material

Vector Group	Words
1	yaz, san, sol, tur, han, kış, kur, hat, men, bit, yay, caz, mil, nur, gel, boz, fan, bel, faz, bek, ver, saz, tay, gür, laz, yar, kel, haz, yas, kıl, dut, dam, vay, yat, pin, güz, kut, kam, poz, zil, naz, gam, fal, çil, kın, kem, yoz, güm
2	yol, bin, dil, çin, kız, ten, baş, hak, dur, kim, taş, kan, yan, sen, dek, gaz, bol, gül, bay, fen, yağ, çay, cam, fas, hal, saf, tim, baz, kol, hit, ham,mor, bal, çam, toz, kas, dal, nem, Şan, yem, kil, şer, bor, cin, zar, tat, sap, mat, ban, sal, hay, tak, sim, koz, zat, kek, tav, lal, hol, gem, pim, sek, hap, şap, ker, kep, kak, şak, fol, sop
3	her, son, tek, kez, hem, var, gün, ben, tam, pek, zor, kaç
4	gol, bar, kır, yaş, has, bas, sur, nam, çan, kaz, yak, jel, yen, çim, kaş, kap, sit, pak, yel, put, sos, tut, doz, vur, sac, bul, zam, çit, tas, hin, şad, lor, keş

5.3. Procedure

In this part of the study, the participants, task, procedure, and stimuli were the same as described in Chapter 3. For each word used in this study, word frequency and sentiment were found as described above. For SII, PESQ and SNR calculations, simultaneously presented target and competing word pairs were determined and analyzed in MATLAB. Derived values can be found in Table A.1 in Appendix A.

5.3.1. Clustering Data into 3 groups

Hearing constitutes a sensory process, whereas listening is a learned skill demanding focused attention and deliberate intention to acquire and effectively utilize the auditory information received. Considering the subjective test conducted in this study, the data can be divided into three groups: 1. “Listened-responded”, 2. “Listened-Not Responded”, 3. “Not listened”.

The first group can be perfectly identified, since the words entered by the participants correctly should have been listened to. In order to determine the other groups and separate the data into three groups, the following procedure was applied.

For each trial, the mean values of SII, SNR, PESQ, Word Frequency, and Pupil Dilation were calculated based on the words participants responded with. Next, these mean values were subtracted from the original values of the words for the specified parameters for each trial where no response was given. Subsequently, a decision was made on which combinations of these parameters, along with pupil dilation, to use in an attempt to identify the words that participants listened to but did not respond to. In doing so, six different tables were created, each incorporating pupil dilation along with a combination of PESQ, SII, SNR, and Word Frequency. Using the k-means clustering method, each table was divided into three groups. This procedure was iterated 10,000 times, investigating whether the responded words could be found and in which cluster they predominantly gathered. Based on the test results, if the words to which participants responded were prevalent in a cluster, that cluster was labeled as the words participants “Listened”. Words present in this cluster but not responded to by participants were labeled as “Listened but not responded”. The data was subsequently categorized into three groups: 1. "Listened-responded", 2. "Listened-No Responded", and 3. "Not listened".

To verify the approach, 70% of the data was randomly selected, and for this selected 70% of the data, the mean values of SII, SNR, PESQ, Word Frequency, and Pupil Dilation for the words to which participants responded were calculated. These mean values were then subtracted from the original values of the words for the specified parameters for each trial where no response was given. The steps described above were followed, and a cluster containing the words that participants responded to as answers was identified. This iteration was repeated 10,000 times, and the results indicated that the selected cluster was the most reliable and suitable. Subsequent analyses were conducted based on this cluster.

5.4. Data Analysis

Firstly, correlation and linear regression analysis conducted using derived attributes presented in Table A.1. In these analyses, we applied correlation and regression methods to determine relationships between attributes and linear modeling of these relationships. Correlation analysis was carried out to explore relationships between variables in the dataset. In cases where the data exhibited a normal distribution, the Pearson correlation test was employed to quantify the strength and direction of the linear relationships between attributes. On the other hand, if the data did not follow a normal distribution, indicating non-linearity or skewness, the Spearman correlation test was utilized.

Later on, we conducted analysis using created categories presented in section 5.4.1. Taking into account the defined dependent and independent variables, linear mixed-effect model analyses were conducted using R (version 4.2.1) with the lme4 package (Bates, Maechker, Bolker, & Walker, 2015). Each dependent variable response was employed as the dependent variable, with fixed effects incorporated for independent variables and their interactions. Participant was incorporated as a random effect. For determining significance, degrees of freedom and generating p-values in the mixed-effect model, the lmerTest package was utilized (Kuznetsova, Brockhoff, & Christensen, 2017). Also, we used the following libraries for analysis; foreign, psych, pastecs, car, reshape2, emmeans, tidyverse, rstatix, afex, ggpubr, ez, multcomp, lme4, lmerTest, ggplot2, sjstats, arm.

These mixed-effect models were employed to assess the effects of the generated categories and other variables on the investigated dependent variables. Considering the effects of categorized variables as well as other significant factors, a more extensive and detailed analysis was performed using these models. These methods enhanced the integrity of our analytical approach by enabling a thorough understanding of the roles of both the attributes and the created categories in the analysis.

5.4.1. Pre-Data Analysis

SII provides a metric between 0 and 1, where;

- SII = 0: Unintelligible, perfect speech privacy
- SII = 0:1: Confidential privacy
- SII = 0:2: Normal privacy
- SII = 0:3: 35% of sentences are correct
- SII = 0:5: 90% of sentences are correct
- SII = 1: Perfectly intelligible

Using above metric, we created three categories to compare SII results.

1- SII= 0.0 - 0.45- “Low”
2- SII= 0.45 - 0.72- “Medium”
3- SII= 0.72- 1.0 - “High”

PESQ score definitions are typically grouped into six bands.

1. 1.00- 1.99 “No meaning understood with any feasible effort”
2. 2.00 - 2.39 “Considerable effort required”
3. 2.40 - 2.79 “Moderate effort required”
4. 2.80 - 3.29 “Attention necessary; a small amount of effort required”
5. 3.30 - 3.79 “Attention necessary; no appreciable effort required”
6. 3.80 - 4.50 “Complete relaxation possible; no effort required.”

Using above metric, we created two categories to compare PESQ results.

1. PESQ= -0.50 – 2.79: “Low Quality”
2. PESQ = 2.80 – 4.5: “High Quality”

We created three categories to compare SNR results.

1. SNR= [-12, -2]
2. SNR= (-2,2]
3. SNR= (2,12]

The log values of the frequency of use of each word were used instead of direct frequency values, because the range of frequencies was large.

1. Word Frequency= [-2, 1]: “Infrequent”
2. Word Frequency= (1, 4]: “Frequent”

5.5. Results

Firstly, we presented correlation and linear regression analysis results using derived attributes. Later on, we presented the results of mixed-effect models using created categories.

5.5.1. Pupil Dilation Correlation and Linear Regression Analysis Results

5.5.1.1. SII Results

Pupil dilation was not significantly correlated with target speech SII values (r = -0.010, p = 0.410). Also, Pupil dilation was not significantly correlated with competing speech SII values (r = -0.011, p = 0.353).

Multiple linear regression was carried out to investigate the relationship between pupil dilation, target speech SII, and competing speech SII. There was not a significant relationship between target speech SII and pupil dilation ($p > 0.05$), competing speech SII and pupil dilation ($p > 0.05$). Also, the interaction of target speech SII and competing speech SII does not have any significant relationship with pupil dilation.

5.5.1.2. PESQ Results

Pupil dilation was significantly correlated with competing speech PESQ values (r = -0.026, p = 0.045). However, Pupil dilation was not significantly correlated with target speech PESQ values (r = -0.019, $p > 0.05$).

Multiple linear regression was carried out to investigate the relationship between pupil dilation, target speech PESQ, and competing speech PESQ. There was a significant relationship between target speech PESQ and pupil dilation (p = 0.005), competing speech PESQ and pupil dilation (p = 0.003) and competing speech PESQ and target speech PESQ interaction (p = 0.014). The adjusted R2 value was 0.002 so 0.2% of the variation in pupil dilation can be explained by the model containing target speech PESQ and competing speech PESQ. The data fulfilled the prerequisites of having homogenous variance and linearity, and the residuals exhibited a close approximation to a normal distribution.

To investigate the relationship with pupil dilation, target speech PESQ, and competing speech PESQ, a linear regression (using lme4 in R) random intercepts model was conducted. Target speech PESQ and competing speech PESQ values were entered as fixed factor and Participant was entered in the model as random effect factor. The relationship between pupil dilation, target speech PESQ and competing speech PESQ showed significant variance in intercepts across participants, SD = 0.192, $\chi 2(2)$ = 13.474, p = 0.001. Pupil dilation significantly predicted target speech PESQ, b = -0.144, t (6045.10) = -3.193, p = 0.001 and significantly predicted competing speech PESQ, b = -0.151, t (6045.10) = -3.309, $p < 0.001$. Also, pupil dilation significantly predicted interaction of target speech PESQ and competing speech PESQ, b = -0.052, t (6045.10) = -2.760, p = 0.006.

5.5.1.3. SNR Results

Pupil dilation was not significantly correlated with target speech SNR values (r = -0.004), and competing Speech SNR value (r = 0.010, $p > 0.05$).

Multiple linear regression was carried out to investigate the relationship between pupil dilation, target speech SNR, and competing speech SNR. There was not a significant relationship between target speech SNR and pupil dilation ($p > 0.05$), competing speech SNR and pupil dilation ($p > 0.05$).

5.5.1.4. Word Frequency Results

Pupil dilation was not significantly correlated with target word frequency values (r = -0.004, $p > 0.05$), and competing word frequency value (r = 0.016, $p > 0.05$).

Multiple linear regression was carried out to investigate the relationship between pupil dilation, target word frequency, and competing word frequency. There was a significant relationship between target word frequency and pupil dilation (p = 0.005) and competing word frequency and target word frequency interaction (p = 0.009). However, there was not a significant relationship between competing word frequency and pupil dilation (p = 0.141). The adjusted R2 value was 0.001 so 0.1% of the

variation in pupil dilation can be explained by the model containing target word frequency and competing word frequency. The data fulfilled the prerequisites of having homogenous variance and linearity, and the residuals exhibited a close approximation to a normal distribution.

To investigate the relationship between pupil dilation and target word frequency, competing word frequency, a linear regression (using lme4 in R) random intercepts model was conducted. Target word frequency and competing word frequency value were entered as fixed factor and Participant was entered in the model as random effect factor. The relationship between pupil dilation, Target word frequency and competing word frequency showed significant variance in intercepts across participants, SD = 0.192, $\chi 2(3) = 11.932$, $p = 0.001$. Pupil dilation significantly predicted target word frequency, $b = -0.028$, t (6045) = -3.078, $p = 0.001$ and significantly predicted competing word frequency, $b = -0.142$, t (6045) = -1.568, $p = 0.001$. Also, pupil dilation significantly predicted interaction of target word frequency and competing word frequency, $b = -1.88$, t (6502) = 2.894, $p < 0.001$.

5.5.1.5. Word Sentiment Results

Pupil dilation was not significantly correlated with target word sentiment values ($r = -0.013$, $p > 0.05$), and competing word sentiment value ($r = -0.022$, $p > 0.05$).

Multiple linear regression was carried out to investigate the relationship between pupil dilation, target word sentiment, and competing word sentiment. There was not a significant relationship between target word sentiment and pupil dilation ($p > 0.05$), competing word sentiment and pupil dilation ($p > 0.05$).

5.5.2. Peak Dilation Correlation and Linear Regression Analysis Results

5.5.2.1. SII Results

Peak dilation was significantly correlated with target speech SII values ($r = -0.015$, $p = 0.020$). However, Peak dilation was not significantly correlated with competing speech SII values ($r = 0.013$, $p > 0.05$).

Multiple linear regression was carried out to investigate the relationship between peak dilation, target speech SII, and competing speech SII. There was not a significant relationship between target speech SII and peak dilation ($p > 0.05$), competing speech SII and peak dilation ($p > 0.05$).

5.5.2.2. PESQ Results

Peak dilation was significantly correlated with target speech PESQ values ($r = -0.026$, $p = 0.026$). However, Peak dilation was not significantly correlated with competing speech PESQ values ($r = -0.009$, $p > 0.05$).

Multiple linear regression was carried out to investigate the relationship between peak dilation, target speech PESQ, and competing speech PESQ. There was a significant relationship between target speech PESQ and peak dilation ($p = 0.031$). However,

competing speech PESQ and peak dilation (p = 0.071) and competing speech PESQ and target speech PESQ interaction (p = 0.067). The adjusted R2 value was 0.0007 so 0.07% of the variation in peak dilation can be explained by the model containing target speech PESQ and competing speech PESQ. The data fulfilled the prerequisites of having homogenous variance and linearity, and the residuals exhibited a close approximation to a normal distribution.

To investigate the relationship between peak dilation and target speech PESQ, competing speech PESQ, a linear regression (using lme4 in R) random intercepts model was conducted. Target Speech PESQ and Competing Speech PESQ value were entered as fixed factor and Participant was entered in the model as random effect factor. The relationship between peak dilation, Target Speech PESQ and Competing Speech showed significant variance in intercepts across participants, SD = 0.248, $\chi 2(3)$ = 11.59, p = 0.009. Peak dilation significantly predicted Target Speech PESQ, b = -0.109, t (6045.05) = -2.753, p = .006 and significantly predicted Competing Speech PESQ, b= -0.092, t (6045.04) = -3.193, p = 0.021. Also, peak dilation significantly predicted the interactions of Target Speech PESQ and Competing Speech PESQ, b = -0.038, t (6045.05) = 2.324, p = 0.020.

5.5.2.3. SNR Results

Peak dilation was not significantly correlated with target speech SNR values (r = -0.008, p > 0.05), and competing Speech SNR value (r = 005, p > 0.05).

Multiple linear regression was carried out to investigate the relationship between peak dilation, target speech SNR, and competing speech SNR. There was not a significant relationship between target speech SNR and peak dilation (p > 0.05), competing speech SNR and peak dilation (p > 0.05).

5.5.2.4. Word Frequency Results

Peak dilation was not significantly correlated with target word frequency values (r = -0.004), and competing word frequency value (r = 0.009, p > 0.05).

Multiple linear regression was carried out to investigate the relationship between pupil dilation, target word frequency, and competing word frequency. There was a not significant relationship between target word frequency and peak dilation (p > 0.05), and competing word frequency and peak dilation (p > 0.05).

5.5.2.5. Word Sentiment Results

Peak dilation was not significantly correlated with target word sentiment values (r = 0.002, p > 0.05) and competing word sentiment value (r = -0.001, p > 0.05).

Multiple linear regression was carried out to investigate the relationship between peak dilation, target word sentiment, and competing word sentiment. There was not a significant relationship between target speech sentiment and peak dilation (p > 0.05), competing speech sentiment and peak dilation (p > 0.05).

To investigate the relationship between peak dilation and target word sentiment, competing word sentiment, a linear regression (using lme4 in R) random intercepts model was conducted. Target word sentiment and Competing Word Sentiment value were entered as fixed factor and Participant was entered in the model as random effect factor. The relationship between peak dilation, Target Word Sentiment and Competing Word Sentiment did not show significant variance in intercepts across participants, *SD* = 0.248, χ2(3) = 6.982, *p* = 0.072. Peak dilation, significantly predicted interactions Target Word Sentiment: Competing Word Sentiment, *b* = -0.332, *t* (6045.06) = -2.282, *p* = .022. However, peak dilation was not significantly predicted Target Word Sentiment, *b* = 0.024, *t* (6045.03) = 0.947, *p* = .344 and significantly predicted Competing Word Sentiment, *b* = -0.024, *t* (6045.05) = -0.904, *p* = .366.

5.5.3. Peak Latency Correlation and Linear Regression Analysis Results

5.5.3.1. SII Results

Peak latency was significantly correlated with target speech SII values (*r* = -0.024, *p* = 0.047). However, Peak latency was not significantly correlated with competing speech SII values (*r* = 0.018, *p* > 0.05).

Multiple linear regression was carried out to investigate the relationship between peak latency, target speech SNR, and competing speech SNR. There was not significant relationship between target speech SNR and peak latency (*p* > 0.05), competing speech SNR and peak latency (*p* > 0.05).

5.5.3.2. PESQ Results

Peak latency was not significantly correlated with target speech PESQ values (*r* = -0.022, *p* > 0.05), and competing speech PESQ values (*r* = -0.015, *p* > 0.05).

Multiple linear regression was carried out to investigate the relationship between peak latency, target speech PESQ, and competing speech PESQ. There was a significant relationship between target speech PESQ and peak latency (*p* = 0.029), and competing speech PESQ and target speech PESQ interaction (*p* = 0.035). However, there was not significant relationship between competing speech PESQ and peak latency (*p* = 0.058). The adjusted R2 value was 0.0005 so 0.05% of the variation in peak latency can be explained by the model containing target speech PESQ and competing speech PESQ. The data fulfilled the prerequisites of having homogenous variance and linearity, and the residuals exhibited a close approximation to a normal distribution.

To investigate the relationship between peak latency and target speech PESQ, competing speech PESQ, a linear regression (using lme4 in R) random intercepts model was conducted. Target Speech PESQ and Competing Speech PESQ value were entered as fixed factor to and Participant was entered in the model as random effect factor. The relationship between peak latency, Target Speech PESQ and Competing Speech PESQ did not show significant variance in intercepts across participants, *SD* = 2501, χ2(3) = 7.26, *p* = 0.064. Peak latency significantly predicted Target Speech PESQ, *b* = -1736.5, *t* (6045.1) = -2.319, *p* = 0.020 and significantly predicted Competing Speech PESQ, *b* = -1495, *t* (6045.1) = -1.980, *p* = 0.048. Also, peak latency

significantly predicted the interactions of Target Speech PESQ and Competing Speech PESQ, $b = 681.5$, t (6045.1) = 2.196, $p = 0.028$.

5.5.3.3. SNR Results

Peak latency was not significantly correlated with target speech SNR values ($r = -0.010$, $p > 0.05$), and competing Speech SNR value ($r = 0.003$, $p > 0.05$)

Multiple linear regression was carried out to investigate the relationship between peak latency, target speech SNR, and competing speech SNR. There was not a significant relationship between target speech SNR and peak latency ($p > 0.05$), competing speech SNR and peak latency ($p > 0.05$).

5.5.3.4. Word Frequency Results

Peak latency was not significantly correlated with target word frequency values ($r = -0.0007$, $p > 0.05$) and competing word frequency value ($r = 0.006$, $p > 0.05$)

Multiple linear regression was carried out to investigate the relationship between peak latency, target word frequency, and competing word frequency. There was a not significant relationship between target word frequency and peak latency ($p > 0.05$), and competing word frequency and peak latency ($p > 0.05$).

5.5.3.5. Word Sentiment Results

Peak latency was not significantly correlated with target word sentiment values ($r = 0.002$, $p > 0.05$) and competing word sentiment value ($r = -0.001$, $p > 0.05$).

Multiple linear regression was carried out to investigate the relationship between peak latency, target word sentiment, and competing word sentiment. There was not a significant relationship between target speech sentiment and peak latency ($p > 0.05$), competing speech sentiment and peak latency ($p > 0.05$).

5.5.4. Accuracy Correlation and Linear Regression Analysis Results

5.5.4.1. SII Results

Overall Accuracy was not significantly correlated with target speech SII values ($r = -0.0006$, $p > 0.05$) and competing speech SII values ($r = 0.003$, $p > 0.05$).

Multiple linear regression was carried out to investigate the relationship between Overall Accuracy, target speech SII, and competing speech SII. There was a significant relationship between target speech SII and Overall Accuracy ($p = 0.029$), competing speech SII and Overall Accuracy ($p = 0.033$) and competing speech SII and target speech SII interaction ($p = 0.017$). The adjusted R2 value was 0.001 so 0.1 % of the variation in Overall Accuracy can be explained by the model containing target speech SII and competing speech SII. The data fulfilled the prerequisites of having homogenous variance and linearity, and the residuals exhibited a close approximation to a normal distribution.

To investigate the relationship between overall accuracy and target speech SII, competing speech SII, a linear regression (using lme4 in R) random intercepts model was conducted. Target Speech SII and Competing Speech SII value were entered as fixed factor and Participant was entered in the model as random effect factor. The relationship between overall accuracy, Target Speech SII and Competing SII showed significant variance in intercepts across participants, $SD = 0.323$, $\chi 2(3) = 12.166$, $p = 0.007$. Overall Accuracy significantly predicted Target Speech SII, $b = 1.148$, $t(6045.18) = 2.382$, $p = 0.017$ and significantly predicted Competing Speech SII, $b = 1.105$, $t(6045.16) = 2.328$, $p = 0.020$. Also, Overall Accuracy significantly predicted the interactions of Target Speech SII and Competing Speech SII, $b = -1.384$, $t(6045.16) = -2.544$, $p = 0.011$.

5.5.4.2. PESQ Results

Overall accuracy was not significantly correlated with target speech PESQ values ($r = 0.020$, $p > 0.05$), but overall accuracy was significantly correlated with competing speech PESQ values ($r = 0.005$, $p < 0.05$).

Multiple linear regression was carried out to investigate the relationship between overall accuracy, target speech PESQ, and competing speech PESQ. There was a significant relationship between target speech PESQ and Overall Accuracy ($p = 0.003$), competing speech PESQ and Overall Accuracy ($p = 0.006$) and competing speech PESQ and target speech PESQ interaction ($p = 0.011$). The adjusted R2 value was 0.001 so 0.1% of the variation in pupil latency can be explained by the model containing target speech PESQ and competing speech PESQ. The data fulfilled the prerequisites of having homogenous variance and linearity, and the residuals exhibited a close approximation to a normal distribution.

To investigate the relationship with overall accuracy and target speech PESQ, and competing speech PESQ, a linear regression (using lme4 in R) random intercepts model conducted. Target Speech PESQ and Competing Speech PESQ values were entered as fixed factor and Participant was entered in the model as random effect factor. The relationship between overall accuracy, Target Speech PESQ and Competing Speech PESQ showed significant variance in intercepts across participants, $SD = 0.248$, $\chi 2(3) = 12.646$, $p = 0.005$. Overall accuracy significantly predicted Target Speech PESQ, $b = 0.309$, $t(6045.13) = 3.042$, $p = 0.002$ and significantly predicted Competing Speech PESQ, $b = 0.289$, $t(6045.12) = 2.817$, $p = 0.005$. Also, overall accuracy significantly predicted the interactions of Target Speech PESQ and Competing Speech PESQ, $b = -0.110$, $t(6045.14) = -2.616$, $p = 0.009$.

5.5.4.3. SNR Results

Overall accuracy was not significantly correlated with target speech SNR values ($r = -0.016$, $p > 0.05$), and competing speech SNR value ($r = 0.015$, $p > 0.05$)

Multiple linear regression was carried out to investigate the relationship between overall accuracy, target speech SNR, and competing speech SNR. There was not a significant relationship between target speech SNR and overall accuracy ($p > 0.05$), competing speech SNR and overall accuracy ($p > 0.05$).

5.5.4.4. Word Frequency Results

Overall accuracy was not significantly correlated with target word frequency values (r = -0.002, p > 0.05), and competing word frequency value (r = -0.003, p > 0.05).

Multiple linear regression was carried out to investigate the relationship between pupil latency, target word frequency, and competing word frequency. There was a not significant relationship between target word frequency and overall accuracy (p >.05), and competing word frequency and overall accuracy (p >.05).

5.5.4.5. Word Sentiment Results

Overall Accuracy was not significantly correlated with target word sentiment values (r = -0.14, p > 0.05), and competing word sentiment value (r = -0.007, p > 0.05).

Multiple linear regression was carried out to investigate the relationship between Overall Accuracy, target word sentiment, and competing word sentiment. There was not a significant relationship between target speech sentiment and Overall Accuracy (p > 0.05), competing speech sentiment and Overall Accuracy (p > 0.05).

5.5.5. Pupil Dilation Linear Mixed Effect Model Results

To examine which factors effect pupil dilation, several linear mixed models were applied (Table 4). Each model was compared with the base model. For the models which significantly improved the prediction of pupil dilation, interaction models were developed.

Table 4: Linear Mixed Effect Model of Pupil Dilation

Model	**χ^2**	**df**	**p**
Base Model : pupil_dilation ~ 1 + (1 | subject)			
model_01: pupil_dilation ~ response_type + (1 | subject)	13.591	2	**.001**
model_02: pupil_dilation ~ p_gender + (1 | subject)	0.002	1	.965
model_03: pupil_dilation ~ s_gender + (1 | subject)	0.013	1	.909
model_04: pupil_dilation ~ age_group + (1 | subject)	0.344	1	.558
model_05: pupil_dilation ~ audio_direction + (1 | subject)	68.802	2	**<.001**
model_06: pupil_dilation ~ target_wordtovec2 + (1 | subject)	9.587	3	**.022**
model_07: pupil_dilation ~ competing_wordtovec2 + (1 | subject)	6.302	3	.098
model_08: pupil_dilation ~ pesq_target + (1 | subject)	10.167	1	**.001**
model_09: pupil_dilation ~ pesq_competing + (1 | subject)	5.758	1	**.016**
model_10: pupil_dilation ~ freq_target + (1 | subject)	1.754	1	.185
model_11: pupil_dilation ~ freq_competing + (1 | subject)	1.651	1	.199
model_12=pupil_dilation ~ positive_negative + (1|subject)	3.510	1	.061
model_13= pupil_dilation ~ sii_target + (1|subject)	1.020	2	.601
model_14= pupil_dilation ~ sii_competing + (1|subject)	2.185	2	.335
model_15= pupil_dilation ~ target_sentiment_category+ (1 | subject)	9820.5	1	**<.001**
model_16= pupil_dilation ~ competing_sentiment_category + (1 | subject)	9823.5	1	**<.001**

Response Type Results

To investigate the relationship between pupil dilation and response type, a linear regression (using lme4 in R) random intercepts model was conducted. Response type value was entered as fixed factor to and Participant was entered in the model as random effect factor. The relationship between the pupil dilation, and the Response type showed significant variance in intercepts across participants, SD = 0.189, $\chi 2(2)$ = 13.591, p = 0.001. Pupil dilation significantly predicted "Listened-Responded", b = -0.041, t (6049) = 3.639 p < 0.001 and significantly predicted "Listened-Not Responded", b = 0.011, t (6049) = 1.043, p = 0.001. Pairwise comparisons using t-tests, corrected with Holm's sequential Bonferroni procedure, indicated that "Not Listened", (M = 0.360, SD = 0.39) vs. "Listened-Responded", (M = 0.401, SD = 0.4) (t (6051) = -3.638, p < 0.001), and "Not-Listened" vs. "Listened-Not Responded" (M = 0.371, SD = 0.39) (t (6051) = 2.622, p = .018) were statistically significantly different but "Listened-Responded" vs. "Listened-Not Responded" (t (6051) = -1.043, p = 0.2971) was not statistically significant.

Figure 21 shows response type effect on pupil dilation.

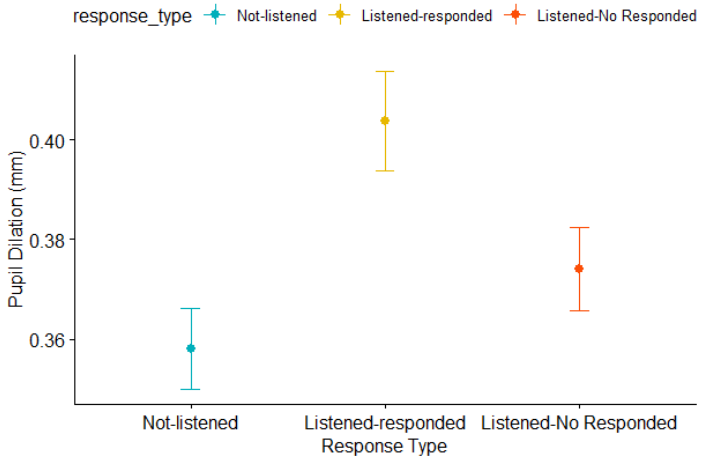


Figure 21: Response type effect on pupil dilation

Audio Direction Results

To investigate the relationship between pupil dilation and audio direction, a linear regression (using lme4 in R) random intercepts model was conducted. Audio Direction value was entered as fixed factor and Participant was entered in the model as random effect factor. The relationship between pupil dilation and, Audio Direction showed significant variance in intercepts across participants, SD = 0.189, $\chi 2(2)$ =68.802, p < 0.001. Pupil dilation significantly predicted "Bilateral Audio Direction", b = 0.055, t (6048.13) = 5.063, p < 0.001 and significantly predicted "Right Audio Direction", b = -0.034, t (6048.15) = -3.212, p = 0.001. Pairwise comparisons using t-tests, corrected with Holm's sequential Bonferroni procedure, indicated that "Left Audio Direction" (M = 0.368, SD = 0.39), vs. "Bilateral Audio Direction" (M = 0.423, SD = 0.4) (t (6050) = -5.062, p < 0.001), "Left Audio Direction" vs. 'Right Audio Direction" (M = 0.334, SD = 0.39) (t (6050) = 3.212, p = 0.001) and 'Bilateral Audio Direction" vs. 'Right

Audio Direction" (t (6051) = 8.255, p < 0.001) were statistically significantly different.

Target Word Vector Results

To investigate the relationship between pupil dilation and target word vector, a linear regression (using lme4 in R) random intercepts model was conducted. Target word vector value was entered as fixed factor and Participant was entered in the model as random effect factor. The relationship between pupil dilation and, target word vector showed significant variance in intercepts across participants, SD = 0.189, $\chi 2(3)$ = 9.587, p = 0.022. Pupil dilation significantly predicted "Second Word Vector Group", b = 0.027, t (6048.10) = 2.597 p = 0.009. Pairwise comparisons using t-tests, corrected with Holm's sequential Bonferroni procedure, indicated that no comparisons were statistically significant.

Figure 22 shows target word vector effect on pupil dilation

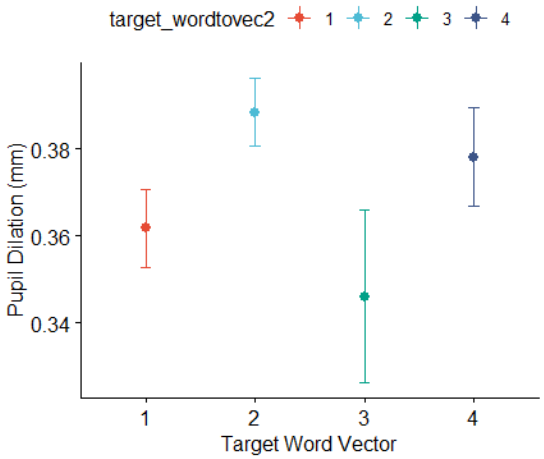


Figure 22: Target word vector effect on pupil dilation

Target Speech PESQ Results

To investigate the relationship between pupil dilation and target speech PESQ, a linear regression (using lme4 in R) random intercepts model was conducted. PESQ target speech value was entered as fixed factor to and Participant was entered in the model as random effect factor. The relationship between pupil dilation and, target speech PESQ showed significant variance in intercepts across participants, SD = 0.189, $\chi 2(1)$ = 10.167, p = 0.001. Pupil dilation significantly predicted High PESQ values, b = -0.036, t (6048.07) = -3.19 p = 0.001. Pairwise comparisons using t-tests, corrected with Holm's sequential Bonferroni procedure, indicated that High PESQ values (M = 0.345, SD = 0.39) and Low PESQ values (M = 0.381, SD = 0.39), (t (6049) = 3.190, p = 0.001) was statistically significantly different.

Figure 23 shows different level of target speech PESQ effect on pupil dilation.

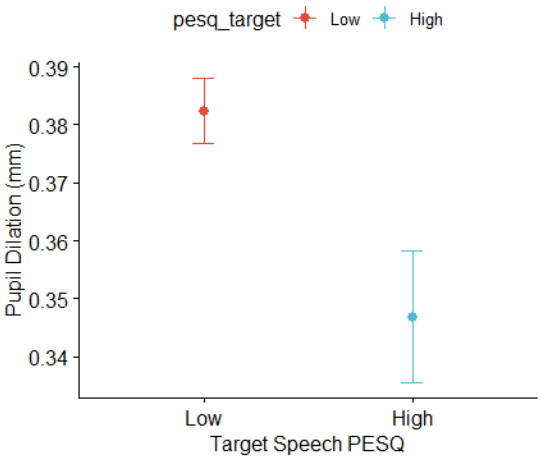


Figure 23: Target word PESQ effect on pupil dilation

Competing Speech PESQ Results

To investigate the relationship between pupil dilation and competing speech PESQ, a linear regression (using lme4 in R) random intercepts model was conducted. PESQ competing speech value was entered as fixed factor and Participant was entered in the model as random effect factor. The relationship between pupil dilation and, competing speech PESQ showed significant variance in intercepts across participants, *SD* = 0.189, $\chi 2(1) = 5.758$, $p = 0.016$. Pupil dilation significantly predicted High PESQ values, $b = -0.026$, t (6048.09) = -2.4 $p = 0.016$. Pairwise comparisons using t-tests, corrected with Holm's sequential Bonferroni procedure, indicated that High PESQ values ($M = 0.354$, $SD = 0.39$), Low PESQ values ($M = 0.380$, $SD = 0.39$) (t (6049) = 2.400, p =.016) was statistically significantly different.

Figure 24 shows different level of competing speech PESQ effect on pupil dilation.

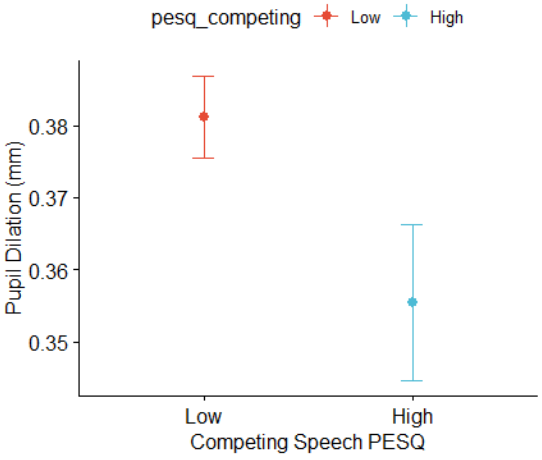


Figure 24: Competing word PESQ effect on Pupil Dilation

5.5.5.1. Interactions Results

Table 5: Interactions of Pupil Dilation Linear Mixed Effect Models

Model	$\chi 2$	df	p
model_00: pupil_dilation ~ 1 + (1 | subject)			
model_01: pupil_dilation ~ response_type + (1 | subject)			
model_02: pupil_dilation ~ response_type * audio_direction + (1 | subject)	97.819	8	**2.2e-16 *****
model_03: pupil_dilation ~ response_type * pesq_target + (1 | subject)	35.094	5	**1.441e-06 *****
model_04: pupil_dilation ~ response_type * pesq_competing + (1 | subject)	21.923	5	**0.0005416 *****
model_05: pupil_dilation ~ response_type * competing_sentiment_category + (1 | subject)	22.047	5	**0.000513 *****
model_06: pupil_dilation ~ response_type * target_wordtovec2 + (1 | subject)	30.04	11	**0.001562 ****

Response Type-Audio Direction Interaction Results

To examine the effect of Audio Direction and Response Type conditions on pupil dilation, a linear mixed-effect model was fitted on participants' pupil dilation, using Audio Direction and Response Type conditions as fixed effect factors, with Participants as random effect factors. There was a significant main effect of Response Type condition $SD = 0.189$, $\chi 2(2) = 7.7672$, $p = 0.020$, effect of Audio Direction $SD = 0.189$, $\chi 2(2) = 37.211$ $p < 0.001$, and interaction between Response Type and Audio Direction $SD = 0.189$, $\chi 2(4) = 12.354$, $p < 0.001$. In "Not-Listened" condition pupil dilation was lower in "Right Audio Direction" ($M = 0.301$, $SD = 0.39$) and "Listened-Responded" condition was higher in "Bilateral Audio Direction" ($M = 0.464$, $SD = 0.39$). Pairwise comparisons using t-tests, corrected with Holm's sequential Bonferroni procedure, 39 test were conducted and indicated that 9 comparisons were statistically significant. In "Not-Listened" conditions, pupil dilation of "Left Audio Direction" ($M = 0.37$, $SD = 0.4$) was bigger than "Right Audio Direction" ($M = 0.301$, $SD = 0.39$) $b = -0.060$, t (6057) = 3.360, $p = 0.020$. Pupil dilation of "Bilateral Audio Direction" ($M = 0.40$, $SD = 0.39$) was bigger than "Right Audio Direction" ($M = 0.301$, $SD = 0.39$) $b = 0.104$, t (6057) = 6.095 $p < 0.001$. In "Listened-Responded" condition, pupil dilation of "Left Audio Direction" ($M = 0.41$, $SD = 0.39$) was bigger than "Right Audio Direction" ($M = 0.34$, $SD = 0.4$) $b = 0.063$, t (6057) = 3.046, $p = 0.049$. Pupil dilation of "Bilateral Audio Direction" ($M = 0.47$, $SD = 0.39$) was bigger than "Right Audio Direction" ($M = 0.34$, $SD = 0.4$) $b = 0.125$, t (6056) =5.917 $p < 0.001$. In "Listened-Not Responded" condition, pupil dilation of "Left Audio Direction" ($M = 0.34$, $SD = 0.4$) was lower than "Bilateral Audio Direction" ($M = 0.42$, $SD = 0.41$) b= -0.062, t (6057) = -3.308, $p = 0.023$.

Response Type- Vector Interaction Results

To examine the effect of Target Word Vector and Response Type conditions on pupil dilation, a linear mixed-effect model was fitted on participants' pupil dilation, using Target Word Vector and Response Type conditions as fixed effect factors, with Participants as random effect factors. There was not significant main effect of Response Type condition $SD = 0.189$, $\chi 2(2) = 3.435$, $p = 0.179$, effect of Target Word

Vector *SD* = 0.189, $\chi 2(3) = 2.196$ $p = 0.533$, and interaction between Response Type and Target Word Vector *SD* = 0.189, $\chi 2(6) = 6.218$, $p = 0.399$.

Response Type- Target Speech PESQ Interaction Results

To examine the effect of Target Speech PESQ and Response Type conditions on pupil dilation, a linear mixed-effect model was fitted on participants' pupil dilation, using Target Speech PESQ and Response Type conditions as fixed effect factors, with Participants as random effect factors. There was a significant main effect of Response Type condition *SD* = 0.189, $\chi 2(2) = 6.727$, $p = 0.035$, effect of Target Speech PESQ *SD* = 0.189, $\chi 2(1) = 21.395$ $p < 0.001$, and interaction between Response Type and Target Speech PESQ *SD* = 0.189, $\chi 2(2) = 11.831$, $p < 0.001$. In "Not-Listened" condition pupil dilation was lower in High PESQ values (*M* = 0.301, *SD* = 0.39) and "Listened-Responded" condition was higher in Low PESQ values (*M* = 0.464, *SD* = 0.39). Pairwise comparisons using t-tests, corrected with Holm's sequential Bonferroni procedure, 15 test were conducted and indicated that 5 comparisons were statistically significant. In "Non-Listened" conditions, pupil dilation of Low PESQ (*M* = 0.378, *SD* = 0.4) was bigger than High PESQ (*M* = 0.301, *SD* = 0.39) $b = 0.777$, $t(6054) = 4.623$, $p < 0.001$.

Response Type- Competing Speech PESQ Interaction Results

To examine the effect of Competing Speech PESQ and Response Type conditions on pupil dilation, a linear mixed-effect model was fitted on participants' pupil dilation, using Competing Speech PESQ and Response Type conditions as fixed effect factors, with Participants as random effect factors. There was a significant main effect of Response Type condition *SD* = 0.189, $\chi 2(2) = 16.028$, $p < 0.001$, but there was not significant effect of Competing Speech PESQ *SD* = 0.189, $\chi 2(1) = 0.334$, $p = 0.855$, and interaction between Response Type and Competing Speech PESQ *SD* = 0.189, $\chi 2(2) = 3.491$, $p = 0.175$. Pairwise comparisons using t-tests, corrected with Holm's sequential Bonferroni procedure, 15 test were conducted and indicated that 2 comparisons were statistically significant.

Response Type: Audio Direction: Target Speech PESQ Interaction Results

To examine the effect of Audio Direction, Target Speech PESQ, and Response Type conditions on pupil dilation, a linear mixed-effect model was fitted on participants' pupil dilation, using Audio Direction, Target Speech PESQ and Response Type conditions as fixed effect factors, with Participants as random effect factors. There was a significant main effect of Response Type condition *SD* = 0.189, $\chi 2(2) = 7.705$, $p = 0.021$, effect of Audio Direction *SD* = 0.189, $\chi 2(2) = 37.980$, $p < 0.001$, and interaction between Response Type and Audio Direction *SD* = 0.189, $\chi 2(4) = 11.539$, $p = 0.021$, but there was not main effect of Target Speech PESQ *SD*= 0.189, $\chi 2(1) = 3.179$, $p = 0.075$. Pairwise comparisons using t-tests, corrected with Holm's sequential Bonferroni procedure, 153 test were conducted and indicated that 34 comparisons were statistically significant.

5.5.6. Peak Dilation Linear Mixed Effect Model Results

To examine which factors, effect peak dilation, several linear mixed models were applied. Each model was compared with the base model (Table 6). For the models which significantly improved the prediction of peak dilation, interaction models were developed.

Table 6: Linear Mixed Effect Model of Peak Dilation

Model	**χ2**	**df**	**p**
Base Model: peak_dilation ~ 1 + (1 | subject)			
model_01: peak_dilation ~ response_type + (1 | subject)	1.107	2	.574
model_02: peak_dilation ~ p_gender + (1 | subject)	0.010	1	.920
model_03: peak_dilation ~ s_gender + (1 | subject)	2.232	1	.135
model_04: peak_dilation ~ age_group + (1 | subject)	0.697	1	.404
model_05: peak_dilation ~ audio_direction + (1 | subject)	130.95	2	**<.001**
model_06: peak_dilation ~ target_wordtovec2 + (1 | subject)	14.179	3	**.003**
model_07: peak_dilation ~ competing_wordtovec2 + (1 | subject)	2.135	3	.545
model_08: peak_dilation ~ pesq_target + (1 | subject)	4.909	1	**.027**
model_09: peak_dilation ~ pesq_competing + (1 | subject)	1.572	1	.210
model_10: peak_dilation ~ freq_target + (1 | subject)	0.040	1	.841
model_11: peak_dilation ~ freq_competing + (1 | subject)	0.708	1	.400
model_12=peak_dilation ~ sii_target+ (1|subject)	8.479	2	**.014**
model_13=peak_dilation ~ sii_competing+ (1|subject)	0.993	2	.609

Audio Direction Results

To investigate the relationship between peak dilation and audio direction, a linear regression (using lme4 in R) random intercepts model was conducted. Audio Direction value was entered as fixed factor and Participant was entered in the model as random effect factor. The relationship between pupil dilation, Audio Direction showed significant variance in intercepts across participants, $SD = 0.245$, $\chi2(2) = 132.38$, $p < 0.001$. Peak dilation significantly predicted "Bilateral Audio Direction", $b = 0.067$, $t(6048) = 7.162$ $p < 0.001$ and significantly predicted "Right Audio Direction", $b = -0.040$, $t(6048) = -4.257$, $p < .001$. Pairwise comparisons using t-tests, corrected with Holm's sequential Bonferroni procedure, indicated that "Left Audio Direction" ($M = 0.942$, $SD = 0.37$), vs. "Bilateral Audio Direction" ($M = 1.009$, $SD = 0.4$) ($t(6050) = -7.161$, $p < 0.001$), "Left Audio Direction" vs. "Right Audio Direction" ($M = 0.902$, $SD = 0.39$) ($t(6050) = 4.256$, $p < 0.001$) and "Bilateral Audio Direction" vs. "Right Audio Direction" ($t(6051) = 11.391$, $p < 0.001$) were statistically significantly different.

Target Word Vector Results

To investigate the relationship between peak dilation and target word vector, a linear regression (using lme4 in R) random intercepts model was conducted. Target word vector value was entered as fixed factor and Participant was entered in the model as random effect factor. The relationship between pupil dilation, Target word vector showed significant variance in intercepts across participants, $SD = 0.245$, $\chi2(3) = 14.179$, $p = 0.003$. Peak dilation significantly predicted Third Word Vector Group, $b = 0.055$, $t(6048) = 2.597$ $p = 0.001$. Pairwise comparisons using t-tests, corrected with Holm's sequential Bonferroni procedure, indicated that Peak Dilation of First Word Vector Group ($M = 0.950$, $SD = 0.39$) was lower than Third Word Vector Group ($M =$

1.005, *SD* = 0.38), *b* = -0.055, *t* (6051) = -3.300, *p* = 0.004. Peak Dilation of Third Word Vector Group (*M* =1.005, *SD* = 0.38) was bigger Forth Word Vector Group (*M* = 0.947, *SD* = 0.38), *b* = 0.059, *t* (6051) = 3.366, *p* = 0.004.

Figure 25 shows different target word vector effect on peak dilation.

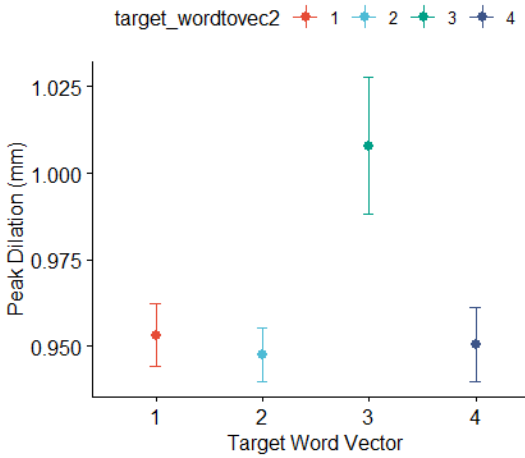


Figure 25: Target word vector effect on peak dilation

Target Speech PESQ Results

To investigate the relationship between peak dilation and PESQ Target, a linear regression (using lme4 in R) random intercepts model was conducted. Target Speech PESQ value was entered as fixed factor and Participant was entered in the model as random effect factor. The relationship between peak dilation, Target Speech PESQ showed significant variance in intercepts across participants, *SD* = 0.245, $\chi 2(1)$ = 4.909, *p* = 0.027. Peak dilation significantly predicted High PESQ values, *b* = -0.022, *t* (6048) = -2.216, *p* = 0.027. Pairwise comparisons using t-tests, corrected with Holm's sequential Bonferroni procedure, indicated that High PESQ values (*M* = 0.94, *SD* = 0.39) and Low PESQ values (*M* = 0.96, *SD* = 0.39), (*t* (6049) = 2.216, *p* = 0.027) was statistically significantly different.

Figure 26 shows different level of competing speech PESQ effect on peak dilation.

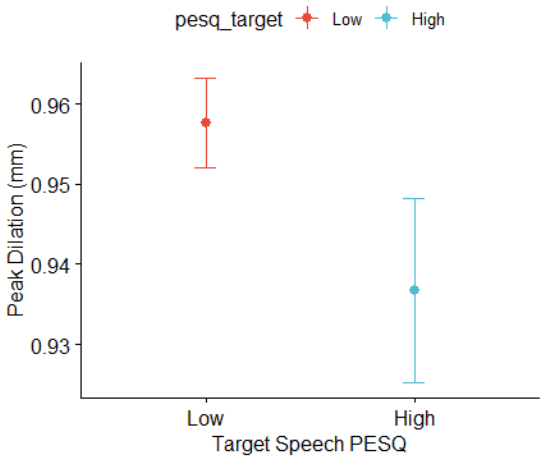


Figure 26: Target word PESQ effect on peak dilation

Target Speech SII Results

To investigate the relationship between peak dilation and Target Speech SII, a linear regression (using lme4 in R) random intercepts model was conducted. Target Speech SII value was entered as fixed factor and Participant was entered in the model as random effect factor. The relationship between peak dilation, Target Speech SII showed significant variance in intercepts across participants, *SD* = 0.245, $\chi 2(2)$ = 8.479, *p* = 0.014. Peak dilation significantly predicted High SII values *b* = -0.079, *t* (6048) = -2.190 *p* = 0.029, and Medium SII Values *b* = -0.017, *t* (6048) = -2.108 *p* = 0.035. Pairwise comparisons using t-tests, corrected with Holm's sequential Bonferroni procedure, indicated that no comparisons were statistically significant.

Figure 27 shows different level of target speech SII effect on peak dilation.

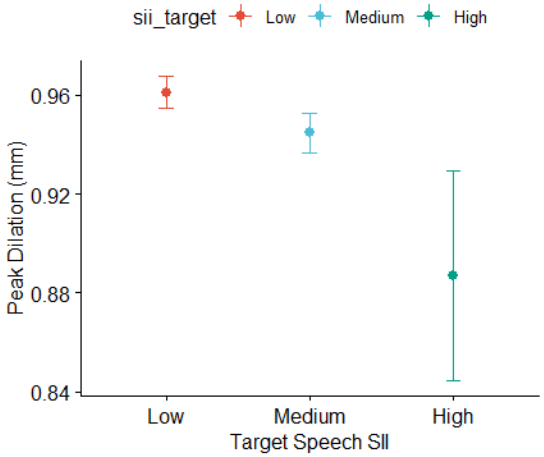


Figure 27: Target word SII effect on peak dilation

5.5.6.1. Interaction Results

Audio Direction-Target word Vector Interaction Results

To examine the effect of Audio Direction, and Target Word Vector conditions on peak dilation, a linear mixed-effect model was fitted on participants' peak dilation, using Audio Direction, and Target Word Vector conditions as fixed effect factors, with Participants as random effect factors. There was a significant main effect of Audio Direction $SD = 0.254$, $\chi2(2) = 37.920$, $p < 0.001$, but there was not main effect of Target Word Vector $SD = 0.254$, $\chi2(3) = 0.109$, $p = 0.991$, and interaction between Target Word Vector and Audio Direction $SD = 0.254$, $\chi2(6) = 4.170$, $p = 0.654$. Pairwise comparisons using t-tests, corrected with Holm's sequential Bonferroni procedure, 66 test were conducted and indicated that 20 comparisons were statistically significant.

Audio Direction- Target Speech PESQ Interaction Results

To examine the effect of Audio Direction, and Target Speech PESQ conditions on peak dilation, a linear mixed-effect model was fitted on participants' peak dilation, using Audio Direction, and Target Speech PESQ conditions as fixed effect factors, with Participants as random effect factors. There was a significant main effect of Audio Direction $SD = 0.254$, $\chi2(2) = 91.633$, $p < 0.001$, but there was not main effect of Target Speech PESQ $SD = 0.254$, $\chi2(1) = 0.067$, $p = 0.796$, and interaction between Target Speech PESQ and Audio Direction $SD = 0.254$, $\chi2(2) = 4.519$, $p = 0.104$. Pairwise comparisons using t-tests, corrected with Holm's sequential Bonferroni procedure, 15 test were conducted and indicated that 10 comparisons were statistically significant.

Target Speech PESQ-Target Word Vector Interaction Results

To examine the effect of Target Speech PESQ, and Target Word Vector conditions on peak dilation, a linear mixed-effect model was fitted on participants' peak dilation, using Target Speech PESQ, and Target Word Vector conditions as fixed effect factors, with Participants as random effect factors. There was a significant main effect of Target Word Vector $SD = 0.254$, $\chi2(3) = 24.695$, $p < 0.001$ and interaction between Target Speech PESQ and Target Word Vector $SD = 0.254$, $\chi2(3) = 12.139$, $p = 0.007$, but there was not main effect of Target Speech PESQ $SD = 0.254$, $\chi2(1) = 0.116$, $p = 0.914$. Pairwise comparisons using t-tests, corrected with Holm's sequential Bonferroni procedure, 28 test were conducted and indicated that 6 comparisons were statistically significant.

5.5.7. Peak Latency Linear Mixed Effect Model Results

To examine which factors, effect peak latency, several linear mixed models were applied. Each model was compared with the base model (Table 7). For the models which significantly improved the prediction of peak latency, interaction models were developed.

Table 7: Linear Mixed Effect Model of Peak Latency

Model	**χ2**	**df**	**p**
Base Model: peak_latency ~ 1 + (1 | subject)			
model_01: peak_latency ~ response_type + (1 | subject)	0.565	2	.753
model_02: peak_latency ~ p_gender + (1 | subject)	1.186	1	.276
model_03: peak_latency ~ s_gender + (1 | subject)	0	1	.999

Model	$\chi 2$	df	p
model_04: peak_latency ~ age_group + (1 | subject)	0.363	1	.547
model_05: peak_latency ~ audio_direction + (1 | subject)	19.617	2	**<.001**
model_06: peak_latency ~ target_wordtovec2 + (1 | subject)	2.817	3	.421
model_07: peak_latency ~ competing_wordtovec2 + (1 | subject)	2.373	3	.499
model_08: peak_latency ~ pesq_target + (1 | subject)	4.013	1	**.046**
model_09: peak_latency ~ pesq_competing + (1 | subject)	1.343	1	.246
model_10: peak_latency ~ freq_target + (1 | subject)	0.197	1	.658
model_11: peak_latency ~ freq_competing + (1 | subject)	0.083	1	.773
model_12=peak_latency ~ positive_negative + (1|subject)	0.001	1	.978
model_13=peak_latency ~ sii_target + (1|subject)	7.586	2	**.023**
model_14=peak_latency ~ sii_competing + (1|subject)	1.066	2	.586
model_15=peak_latency ~ sii_target2 + (1|subject)	1.9911	1	.158
model_16=peak_latency ~ sii_competing2 + (1|subject)	1.066	1	.302
model_17=peak_latency ~ target_sentiment_category + (1 | subject)	8e-04	1	.978
Model=18=peak_latency ~ competing_sentiment_category + (1 | subject)	0.1033	1	.7479

Audio Direction Results

To investigate the relationship between peak latency and audio direction, a linear regression (using lme4 in R) random intercepts model was conducted. Audio Direction value was entered as fixed factor and Participant was entered in the model as random effect factor. The relationship between peak latency, Audio Direction showed significant variance in intercepts across participants, SD = 2471, $\chi 2(2) = 19.617$, $p < 0.001$. Peak dilation significantly predicted "Bilateral Audio Direction", b = 532.6, t (6048.1) = 2.981, p = 0.003. Pairwise comparisons using t-tests, corrected with Holm's sequential Bonferroni procedure, indicated that "Left Audio Direction" (M = 11154, SD = 6440.14), vs. "Bilateral Audio Direction" (M = 11687, SD = 5706.58) (t (6050) = -2.980, p = 0.006), and "Bilateral Audio Direction" vs. "Right Audio Direction" (t (6050) = 4.339, p < 0.001) were statistically significantly different but "Left Audio Direction" vs. "Right Audio Direction" (M = 10912, SD = 6369.59) (t (6050) = 1.367, p = 0.172) was not statistically significantly different.

Target Speech PESQ Results

To investigate the relationship between peak dilation and PESQ Target, a linear regression (using lme4 in R) random intercepts model was conducted. Target Speech PESQ value was entered as fixed factor and Participant was entered in the model as random effect factor. The relationship between peak dilation, Target Speech PESQ showed significant variance in intercepts across participants, SD = 0.245, $\chi 2(1) = 4.909$, p = 0.027. Peak latency significantly predicted High PESQ values, b = -371.65, t (6048.02) = -2.004, p = 0.045. Pairwise comparisons using t-tests, corrected with Holm's sequential Bonferroni procedure, indicated that High PESQ values (M = 10946, SD = 6287.28) and Low PESQ values (M = 11317, SD = 6117.29), (t (6049) = 2.003, p = 0.045) was statistically significantly different.

Figure 28 shows different level of target speech PESQ effect on peak latency.

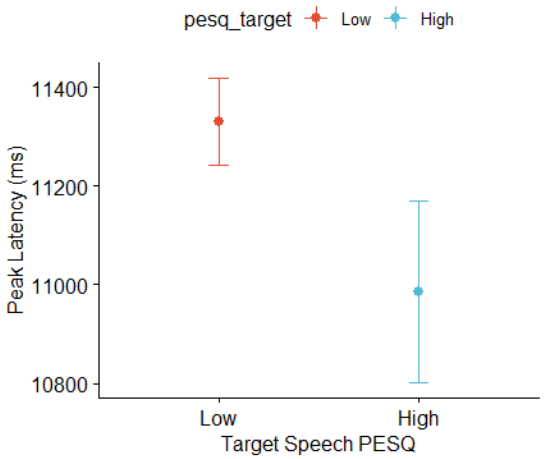


Figure 28: Target word PESQ effect on peak latency

Target Speech SII Results

To investigate the relationship between peak latency and Target Speech SII, a linear regression (using lme4 in R) random intercepts model was conducted. Target Speech SII value was entered as fixed factor and Participant was entered in the model as random effect factor. The relationship between peak latency, Target Speech SII showed significant variance in intercepts across participants, $SD = 2467$, $\chi 2(2) = 7.586$, p = 0.023. Peak latency significantly predicted Medium SII Values b = -338.20, t (6048.07) = -2.266, p = 0.024. Pairwise comparisons using t-tests, corrected with Holm's sequential Bonferroni procedure, indicated that no comparisons were statistically significant.

Figure 29 shows different level of target speech SII effect on peak latency.

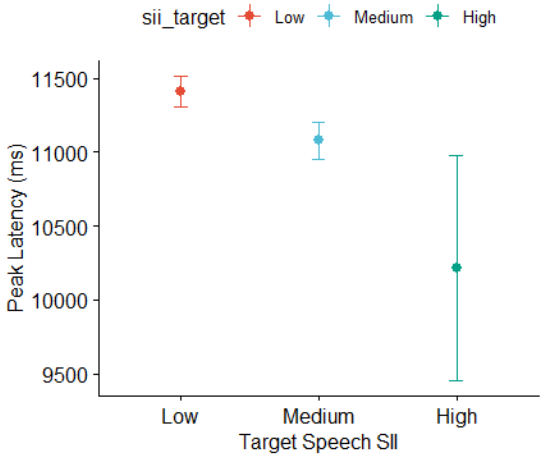


Figure 29: Target word SII effect on peak latency

5.5.6.2. Interactions

Audio Direction- Target Speech PESQ

To examine the effect of Audio Direction, and Target Speech PESQ conditions on peak dilation, a linear mixed-effect model was fitted on participants' peak dilation, using Audio Direction, and Target Speech PESQ conditions as fixed effect factors, with Participants as random effect factors. There was a significant main effect of Audio Direction $SD = 0.254$, $\chi2(2) = 8.938$ $p = 0.011$, but there was not main effect of Target Speech PESQ $SD = 0.254$, $\chi2(1) = 0.840$ $p = 0.259$, and interaction between Target Speech PESQ and Audio Direction $SD= 0.254$, $\chi2(2) = 4.822$, $p = 0.090$. Pairwise comparisons using t-tests, corrected with Holm's sequential Bonferroni procedure, 15 test were conducted and indicated that only 2 comparisons were statistically significant.

5.5.8. Overall Accuracy Linear Mixed Effect Model Results

To examine which factors, effect overall accuracy, several linear mixed models were applied. Each model was compared with the base model (Table 8). For the models which significantly improved the prediction of overall accuracy, interaction models were developed.

Table 8: Linear Mixed Effect Models of Overall Accuracy

Model	**χ2**	**df**	**p**
Base Model: overall_accuracy ~ 1 + (1 | subject)			
model_01: overall_accuracy ~ response_type + (1 | subject)	80.032	2	**<.001**
model_02: overall_accuracy ~ p_gender + (1 | subject)	0.447	1	.504
model_03: overall_accuracy ~ s_gender + (1 | subject)	15.84	1	**<.001**
model_04: overall_accuracy ~ age_group + (1 | subject)	0.018	1	.893
model_05: overall_accuracy ~ audio_direction + (1 | subject)	75.58	2	**<.001**
model_06: overall_accuracy ~ target_wordtovec2 + (1 | subject)	22.542	3	**<.001**
model_07: overall_accuracy ~ competing_wordtovec2 + (1 | subject)	4.683	3	.197
model_08: overall_accuracy ~ pesq_target + (1 | subject)	2.515	1	**.**113
model_09: overall_accuracy ~ pesq_competing + (1 | subject)	0.883	1	.348
model_10: overall_accuracy ~ freq_target + (1 | subject)	0.047	1	.829
model_11: overall_accuracy ~ freq_competing + (1 | subject)	0.308	1	.579
model_12=overall_accuracy ~ sii_target + (1|subject)	10.813	2	**.005**
model_13=overall_accuracy ~ sii_competing + (1|subject)	2.591	2	.274
model_14=overall_accuracy ~ sii_target2 + (1 | subject)	0.015	1	0.902
model_15: overall_accuracy ~ sii_competing2 + (1 | subject)	1.67	1	0.1963
model_16: overall_accuracy ~ target_sentiment_category + (1 | subject)	0.167	1	0.683
model_17: overall_accuracy ~ competing_sentiment_category + (1 | subject)	0.259	1	0.611

Response Type Results

To investigate the relationship between overall accuracy and response type a linear regression (using lme4 in R) random intercepts model was conducted. Response type value was entered as fixed factor and Participant was entered in the model as random effect factor. The relationship between pupil dilation, and Response Type showed significant variance in intercepts across participants, $SD = 0.316$, $\chi2(2) = 80.032$, $p < 0.001$. Overall Accuracy significantly predicted "Listened-Responded", $b = 0.022$, $t(6050) = 8.631$ $p < 0.001$. Pairwise comparisons using t-tests, corrected with Holm's sequential Bonferroni procedure, indicated that "Not-Listened", ($M = 2.35$, $SD = 0.86$) vs. "Listened-Responded", ($M = 2.56$, $SD = 0.73$) $b = -0.215$ ($t(6052) = -8.630$,

p<.001), and "Listened-Responded" vs. "Listened-Not Responded" (M = 2.38, SD = 0.87) b = 0.180 (t (6051) = 2.622, p < 0.001) were statistically significantly different but "Not-Listened" vs. "Listened-Not Responded" b = -0.215 (t (6051) = -1.043, p = 0.2971) was not statistically significant.

Audio Direction Results

To investigate the relationship between overall accuracy and audio direction, a linear regression (using lme4 in R) random intercepts model was conducted. Audio Direction value was entered as fixed factor and Participant was entered in the model as random effect factor. The relationship between overall accuracy, and Audio Direction showed significant variance in intercepts across participants, SD = 2471, $\chi 2(2) = 19.617$, p < 0.001. Overall Accuracy significantly predicted "Bilateral Audio Direction", b = -0.145, t (6048.19) = -6.016, p < 0.001 and "Right Audio Direction", b = 0.060, t (6048.22) = 2.490 p = 0.013. Pairwise comparisons using t-tests, corrected with Holm's sequential Bonferroni procedure, indicated that "Left Audio Direction" (M = 2.44, SD = 0.88), vs. "Bilateral Audio Direction" (M = 2.30, SD = 0.84) b = 0.145 (t (6050) = 6.015, p < 0.001), "Bilateral Audio Direction" vs. "Right Audio Direction" (M = 2.50, SD = 0.77) b = -0.205 (t (6050) = 8.491, p < 0.001) and "Left Audio Direction" vs. 'Right Audio Direction" b =-0.060 (t (6050) = -2.490, p = 0.013) were statistically significantly different.

Speaker Gender Results

To investigate the relationship between overall accuracy and speaker gender, a linear regression (using lme4 in R) random intercepts model was conducted. Speaker Gender was entered as fixed factor and Participant was entered in the model as random effect factor. The relationship between overall accuracy, and Speaker Gender showed significant variance in intercepts across participants, SD = 0.319, $\chi 2(1) = 15.84$, p < 0.001. Overall Accuracy significantly predicted Female Speaker, b = -0.079, t (6048.34) = -3.983, p < 0.001. Pairwise comparisons using t-tests, corrected with Holm's sequential Bonferroni procedure, indicated that Overall Accuracy in the case of Male Speaker (M = 2.46, SD = 0.83), was statistically significantly different than in the case of Female Speaker (M = 2.38, SD = 0.84) b = 0.789 (t (6049) = 3.982, p < 0.001), was statistically significantly different.

Target Word Vector Results

To investigate the relationship between overall accuracy and target word vector, a linear regression (using lme4 in R) random intercepts model was conducted. Target word vector value was entered as fixed factor and Participant was entered in the model as random effect factor. The relationship between overall accuracy, and Target word vector showed significant variance in intercepts across participants, SD = 0.319, $\chi 2(3) = 22.542$, p < 0.001. Overall Accuracy significantly predicted Word Vector Third Group, b = -0.149, t (6048.20) = -3.474, p < 0.001. Pairwise comparisons using t-tests, corrected with Holm's sequential Bonferroni procedure, indicated that Overall Accuracy of Word Vector First Group (M = 2.43, SD = 0.84) was higher than Word Vector Third Group (M = 2.28, SD = 0.85), b = 0.149, t (6051) = 3.473, p = 0.004. Overall Accuracy of Word Vector Second Group (M = 2.40, SD = 0.83) was higher

than Word Vector Third Group (*M* = 2.28, *SD* = 0.85), *b* = 0.118, *t* (6051) = 2.818, *p* = 0.015 and Word Vector Third Group was smaller than Word Vector Forth Group (*M* = 2.48, *SD* = 0.83), *b*= -0.199, *t* (6051) = -4.462, *p* < 0.001.

Figure 30 shows different target word vector effect on overall accuracy.

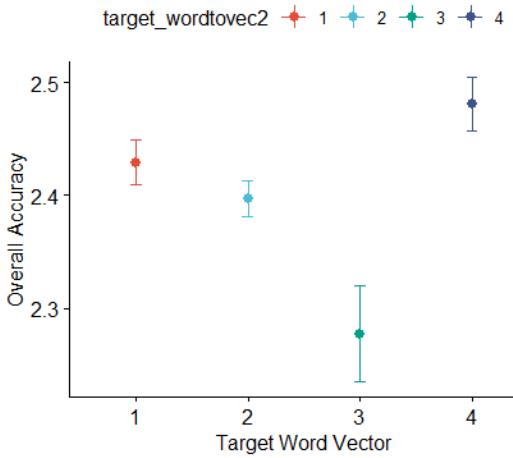


Figure 30: Target word vector effect on overall accuracy

Target Speech SII

To investigate the relationship between Overall Accuracy and Target Speech SII, a linear regression (using lme4 in R) random intercepts model was conducted. Target Speech SII value was entered as fixed factor and Participant was entered in the model as random effect factor. The relationship between Overall Accuracy, and Target Speech SII showed significant variance in intercepts across participants, *SD* = 0.319, $\chi 2(2) = 10.813$, *p* = 0.005. Overall Accuracy significantly predicted High SII Values *b* = -0.303 *t* (6048) = 3.280, *p* = 0.001. Pairwise comparisons using t-tests, corrected with Holm's sequential Bonferroni procedure, indicated that Overall Accuracy in the case of Low SII values (*M* = 2.41, *SD* = 0.83) was significantly smaller than in the case of High SII values (*M* = 2.72, *SD* = 0.7) *b* = -0.303, *t* (6050) = -3.279, *p* = 0.003 and Medium SII values (*M* = 2.42, *SD* = 0.84) was significantly smaller than in the case of High SII values. However, Overall Accuracy was similar in the case of Low SII and Medium SII.

Figure 31 shows different level of target speech SII effect on overall accuracy.

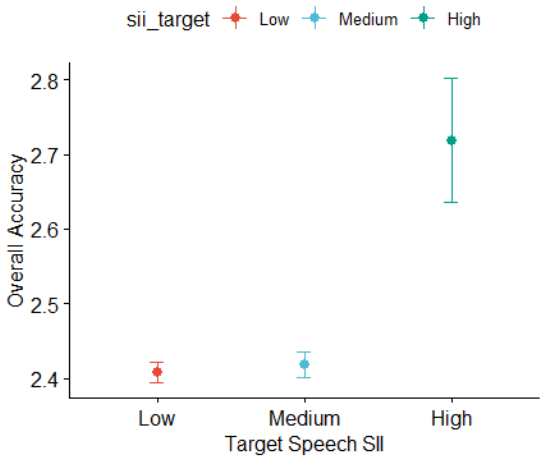


Figure 31: Target word SII effect on overall accuracy

5.6. Discussion

The purpose of this study was to investigate the intelligibility of the words played to the participants, the effect of the semantic content of these words on pupil size, and the difference between hearing and listening in terms of pupil dilation. Specifically, we aimed to understand the relationship between pupil dilation and various measures of speech intelligibility, including SII, PESQ, and SNR values. Additionally, we analyzed the stimulus material in terms of semantics to determine if there was a relationship between the frequency of the words, vector analysis, sentiment analysis, and pupil dilation.

To summarize the main findings, participants' pupil dilation increased as speech intelligibility decreased (as indicated by lower PESQ and SII values). There was a strong relationship between these measures and pupil dilation. However, no significant effect of SNR value on pupil size has been detected, possibly due to the use of continuous SNR values. The content analysis of the words revealed that their popularity or emotional meaning (positive or negative) had no effect on pupil dilation, but vector analysis of the words did have an effect. This may be the result of the participants trying to complete the words in a semantic unity (sentence). The results also showed that hearing and listening had different effects on pupil dilation; when focused on and actively listening to a speech, pupil size increased.

5.6.1. Effects of Speech Intelligibility Metrics

Pupil response consistently increases as the intelligibility of presented speech decreases, unless the intelligibility is extremely low (Zekveld and Kramer, 2014; Zekveld et al., 2010). Results showed that the Perceptual Speech Quality (PESQ) of a speech can affect pupil dilation. When listening to high-quality words, smaller pupil dilation occurs compared to when listening to low-quality words. In other words, pupil dilation is greater for poor sound quality. As Loizou (2011) noted, higher SII and PESQ values, which indicate higher speech intelligibility, resulted in smaller pupil dilation. The PESQ value of a speech can be used to predict changes in pupil dilation

and peak dilation. Peak dilation is larger when listeners focus on low-quality speech, and peak latency is faster for high-quality PESQ values. In other words, peak latency is higher for low-quality words. Additionally, PESQ value affects participants' performance, as those listening to high-quality words had higher overall accuracy.

The results of this study support the literature that PESQ is a measure of speech intelligibility (Liu, 2006) and can be used in studies investigating the relationship between speech intelligibility and pupil dilation (Ma et al., 2009). There is a strong correlation between PESQ scores and speech intelligibility ratings in certain situations, ranging from 0.91 to 0.99 (Beerends et al., 2004; 2005; 2009). Pupil dilation can be measured in studies of listening attention or effort by using stimuli with different PESQ values. This study used Turkish speech as the stimulus material and found that PESQ was a useful predictor of speech intelligibility in Turkish, as indicated by the pupil dilation results. There is debate in the literature about whether PESQ is language dependent; some studies have found it to be language dependent (F. Ben Ali; Chong, 2005), while others have found no significant difference between languages and concluded that PESQ is not language dependent (Triyason & Kanthamanon, 2012).

SII, another metric used in the study, is a measure of speech intelligibility for adults under realistic conditions, as reported by Cahill (2010), Beutelmann (2006), Rhebergen (2005), and Lee (2019). However, SII values were not found to be associated with pupil dilation values in these studies. This may be because, as Brand (2008) noted, different types of spoken material may require different processing methods to produce the clearest and most understandable output. In general, more extensive processing is needed for material with context, such as sentences, compared to single words, as the additional words can provide context that helps the listener understand the intended meaning. Since this study used single words, the results may not accurately reflect speech intelligibility and thus may not have shown a relationship with pupil dilation. Future studies could investigate the relationship between pupil dilation and SII values obtained using sentences instead of single words. It was found that peak dilation decreased as the degree of intelligibility of speech increased. The relationship between target SII and peak dilation was determined, and peak dilation can be predicted using the SII value at high and medium values of target SII, i.e., when the speech is understandable. In addition, the relationship between target speech SII and peak latency was determined, and peak latency can be predicted using moderate SII values. When SII was low, peak latency was also found to be high. Moreover, it was determined that participant performance can be predicted using high SII values; participants performed better at higher SII values.

Existing literature contains studies indicating that Speech Intelligibility Index (SII) exhibits language-dependency. For instance, Galburn and Kitapci (2016) investigated the influence of room acoustics on speech intelligibility across four languages (English, Polish, Arabic, and Mandarin), and their findings revealed noteworthy variations among languages in both listening and predictive speech intelligibility test outcomes. Furthermore, Jin (2014) stated that two key components of SII, BIF and BAF, are language-dependent metrics, making SII itself a language-dependent metric that should be developed specifically for each language. The frequency band importance functions used in SII calculation also vary depending on various factors, including the speech spectrum of the speaker, the language, and the phonemic content

of the stimuli. As a result, SII may not be an effective tool for measuring speech intelligibility in Turkish using pupil dilation due to its language dependency.

Contrary to what has been reported in the literature, no correlation was found between SNR values and pupil dilation in this study. This may be due to the use of continuous SNR values, which differs from the use of specific SNR values in other studies.

5.6.2. Word Frequency and Semantic Content Results

According to the results of the semantic content, or the popularity of the word being played, the frequency of the word affects pupil dilation. As Chapman and Hallowell (2015) and Elshtain and Schaefer (1968) noted, participants who listened to low-frequency words that are semantically difficult had larger pupil dilation than those who focused on high-frequency words. Similarly, we found that participants who listened to low-frequency words had larger pupil dilation, but this difference was not statistically significant. Our results supported those of Papesh and Goldinger (2012). The pupil dilation can be predicted based on the frequency of the word being used. Contrary to Kuchinke et al. (2007), we did not find a relationship between the frequency of the words and peak dilation, nor was a significant relationship found between peak latency and overall accuracy. On the other hand, unlike Kuchinke et al. (2007) and Partala and Surakka (2003), we found that the emotional nature of the words was related to pupil dilation. In particular, larger pupil dilation occurred when the competing target words were positive, but this was not observed for negative words. There was no relationship between overall accuracy and the emotional valence of the words. Similarly, we found that the emotional valence of the words had a relationship with peak dilation, but, contrary to Challis and Krane (1988) and Kuchinke et al. (2005), the positive or negative nature of the words had no effect on participants' overall accuracy or peak latency. Although some studies have discussed the effect of the content of the stimuli, specifically the sentence complexity, on pupil dilation and intelligibility (Kramer et al., 2013; Piquado et al., 2010; Wendt et al., 2016), we were unable to clearly demonstrate this effect in our study due to using only three-letter words. Since the words in the stimulus played were difficult enough, the difference between the words was not clearly observable.

According to the Word2Vector values of another study we conducted on semantic content, the vector values of the words played in the target stimulus are related to pupil dilation. In particular, pupil dilation can be predicted for participants who listen to the words in the second vector group. These participants had larger pupil dilation. Similarly, the vector values of the words also affect peak dilation. Peak dilation can be predicted, especially for participants who listen to the words in the third vector group, using their vector values. These participants had larger peak dilation compared to those who listened to other words. Although peak latency was found to be higher when listening to the words in the third vector group, there was no statistically significant result for the vector values of the target words. However, there was no significant relationship between the vector values of the words and peak latency. Additionally, the words in this group also affected the performance of the participants, with fewer responses being preferred. In addition, the vector values of the words also affected overall accuracy. It is possible to predict overall accuracy using the words in the fourth

vector group. It was also found that the words in the fourth vector group were preferred more as responses by the participants.

5.6.3. Hearing or Listening Effects

In this study, we analyzed the duration of each word separately that we played to the participants. We divided the listening moments into three clusters according to the response given by the participants. We found that the results for the moments when the participants listened and wrote a response, listened but did not write a response, and did not listen at all were different. According to the results, pupil dilation can be predicted during the listening moments. There is a significant difference between the listening moments and the non-listening moments. The pupils enlarge more during the moments when we listen than during the moments when we do not listen. In addition, it has been found that the pupils are larger during the moments when we listen and write a response. However, there is no difference in peak dilation and peak latency among the three groups.

The results of the first study also show that the direction of the sound has an effect on pupil dilation. The results show that pupil dilation can be predicted with the values of the sound coming from the right side and from both sides. Similarly, the direction of the sound has an effect on peak dilation. The peak dilation can be predicted with the values of the sound coming from the right side and from both sides. In addition, the direction of the sound has an effect on peak latency. The peak latency can be predicted with the values of the sound coming from both sides. The performance results also show the effect of audio direction. The overall accuracy can be predicted with the values of the sound coming from the right side and from both sides.

It has been determined that the gender of the speaker affects their performance. The accuracy can be predicted if the speaker is female. It has been found that the participants were more successful when the speaker was male (see Chapter 4).

5.7. Session Conclusion

It has been determined that words with different sound qualities affect pupil dilation. Pupil dilation, peak dilation, and peak latency are larger at low PESQ values. And participants who listen to words with higher quality have higher performance. The relationship between PESQ values and pupil dilation has been established for Turkish. If research is conducted in different languages on the relationship between PESQ values and pupil dilation, findings that support our results can be obtained. Although no effect of SII on monosyllabic words was found, when wider ranges were examined, it was found to be effective on peak dilation and peak latency. The relationship between pupil dilation, peak dilation, peak latency, and overall accuracy and the popularity of words was not found. The positive or negative meaning of the words only affected pupil dilation; no relationship was found with other results. The vector values of the words also had an effect on pupil dilation, which was determined. New studies to support this can be conducted. And the results of this study show that hearing and listening are different concepts and listeners' pupils can show this difference. Pupils are larger when we listen than when we do not listen.

CHAPTER 6

GENERAL CONCLUSIONS

The overarching aim of this comprehensive study was to unveil the complex intricacies of auditory attention by deciphering its directional dynamics and semantic nuances through the pupil responses. With a dual-fold focus on directional auditory attention and the interplay of speech intelligibility metrics and semantic content, this investigation has yielded profound insights into the interplay between auditory processing, cognitive engagement, and linguistic content.

Through rigorous data analysis and interpretation, we have endeavored to provide insightful responses to each of these research queries, paving the way for a comprehensive understanding of the nuanced mechanisms that underlie auditory perception and attention.

Research Question 1: *Does the level of auditory attention have a relation with pupil size and eye movement in processing speech?*

We uncovered a noteworthy correlation between auditory attention level and pupil size during speech stimuli processing. As auditory attention heightened, there was a corresponding dilation of the pupils. This aligns with prior research by Kahneman & Beatty (1966) and Zekveld (2018), reinforcing the intricate connection between cognitive auditory attention processes and physiological responses, as evidenced by pupillometry. These findings underscore the potential of pupil size as a consistent and reliable indicator of auditory attention, consolidating the significance of utilizing pupillometry in understanding cognitive processes during auditory tasks

Research Question 2: *Does interference have an effect on auditory attention, and can this effect be determined through changes in pupil size?*

The study highlighted a significant finding regarding multi-talker speech and its impact on pupil dilation. Interference presence notably influenced pupil dilation, with participants displaying larger pupils during interference, indicating increased cognitive load in managing competing auditory inputs. This aligns with previous research (Koelewijn et al., 2012, 2014; Zekveld et al., 2014, 2018), reaffirming larger pupil sizes, pupil dilations, and peak pupil sizes in the presence of a competing speaker. Moreover, heightened pupil dilations were observed when attention was directed towards the competing speaker rather than the target speaker.

Research Question 3: *Does listener gender affect auditory attention?*

The gender of the participants emerged as a silent influence, leading to changes in how they paid attention to sounds. Men demonstrated a preference for directing their gaze to the right, while women exhibited a tendency for leftward shifts. An exploration into baseline pupil sizes revealed distinct gender-related patterns, highlighting the

complexity of gender-related dynamics in auditory processing. In addition, females exhibited larger pupil sizes overall, but males had greater pupil dilations.

The orchestration of eye movements composed an image of age-driven transformations, unveiling distinct inclinations among younger and middle-aged participants. This interwove with patterns of pupil dilation, emphasizing the intricate interplay between age and auditory attention. The younger participants exhibited more significant pupil dilations, in line with Piquado et al.'s (2010) observation regarding the age-related relationship with a smaller dynamic pupil range. Conversely, older adults demonstrated smaller pupil dilations, consistent with Koch and Janse's (2016) assertion that pupil dilation varies with age, generally being more pronounced in younger individuals.

Research Question 4: *Does the source location of the sound affect listening to a task?*

It emerged that individuals exhibit heightened attentiveness and enhanced task performance when focusing their auditory faculties on stimuli presented to a single ear, rather than splitting their attention between both ears. When participants heard sounds from both sides, their pupils exhibited an increase in size. Particularly during bilateral listening, the interference led to a substantial dilation of their pupils, resulting in significantly larger peak pupil sizes.

The participants consistently directed their eye movements toward the sound source, aligning with Hendrikse et al.'s (2018) and Vertegaal et al.'s (2001) prior findings. Whether the sound originated from the left or right, participants' eyes moved in the corresponding direction, mirroring their auditory focus. This aligns with previous research demonstrating a rapid and predominant shift in gaze towards the speaker, validating the robustness and replicability of this eye movement pattern.

Research Question 5*: Can we investigate the difference between hearing and listening using eye movements?*

Through careful examination of participants' eye movements during auditory processing tasks, we shed light on the nuances that distinguish mere hearing from active listening. The patterns of eye movements we observed, particularly the focus and shifts in gaze, indicated that listening involves a dynamic cognitive process that extends beyond the auditory reception of sounds. This suggests that eye movement analysis can indeed serve as a valuable tool to explore and elucidate the cognitive complexities inherent in listening, providing new dimensions to our understanding of auditory perception and cognition.

Research Question 6: *What is the effect of SII (Speech Intelligibility Index), PESQ (Perceptual Evaluation of Speech Quality), and SNR (Signal-to-Noise Ratio) values on auditory attention and pupil responses?*

It is evident that distinct sound qualities lead to notable differences in pupil dilation, peak dilation, and peak latency, particularly evident at low PESQ values. Furthermore, a noteworthy correlation has emerged between the quality of words listened to and participants' overall performance. This corresponds to the situation as indicated in the literature. Pupil response consistently increases as the intelligibility of the presented speech decreases, unless the intelligibility is extremely low (Zekveld and Kramer, 2014; Zekveld et al., 2010). Particularly insightful is the establishment of a robust relationship between PESQ values and pupil dilation, a linkage validated within the context of the Turkish language. Although Cahill (2010), Beutelmann (2006), Rhebergen (2005), and Lee (2019) highlighted the reliability of the Speech Intelligibility Index (SII) as a measure of speech intelligibility, this present study did not find a distinct correlation between SII values and auditory attention or pupil dilation. Notably, while monosyllabic words showed no discernible impact of SII, a more comprehensive examination revealed its influence on peak dilation and peak latency. Contrary to what has been reported in the literature, no correlation was found between SNR values and pupil dilation in this study

Research Question 7: *Does the semantic content have an impact on auditory attention, and how does this impact manifest in pupil dilation?*

Semantic content significantly influences auditory attention, evidenced by the dilation of pupils. Words, extending beyond phonetic structures, intricately convey meaning and evoke diverse pupil reactions. This intricate relationship between word meaning and pupil responses highlights the fusion of linguistic and physiological aspects in cognitive processing. Moreover, emotional valence also plays a crucial role, as positive words trigger more pronounced pupil responses, illustrating the fusion of affective dimensions with cognitive engagement. Interestingly, in contrast to Kuchinke et al. (2007) and Partala and Surakka (2003), we observed a distinct link between the emotional nature of words and pupil dilation. Specifically, larger pupil dilation was evident when the competing target words carried a positive emotional tone, whereas this effect was not observed for negative words.

Research Question 8: *How does the familiarity of stimuli presented within the study influence auditory attention and pupil dilation?*

Consistent with Chapman and Hallowell (2015) and Elshtain and Schaefer (1968), our study revealed that participants engaged with semantically challenging, low-frequency words exhibited increased pupil dilation compared to those focusing on high-frequency words. While our findings aligned with this trend, the difference in pupil dilation for low-frequency words, though larger, did not reach statistical significance. This concurs with the results reported by Papesh and Goldinger (2012).

In conclusion, this study gives us a clear picture of how our attention to sounds and our pupils' reactions are connected. By understanding this connection, we can come up with new ways to help people who might have trouble with their hearing or focus. Moreover, this study examines whether presented stimuli involve familiar or unfamiliar words, probing emotional connotations effects on pupil dilation responses.

The research establishes a link between speech quality and pupil dilation, indicating better performance with higher-quality auditory stimuli, regardless of lexical popularity or meaning. Ultimately, the study underscores the distinction between hearing and listening, highlighting pupil responses as a window into these cognitive processes.

6.1. Practical Implications

The implications of this study for practical applications are extensive and align with the expanding field of eye movement and pupillometry applications highlighted by Zekveld et al. (2018) and Van Engen & McLaughlin (2018). Leveraging eye movement and pupillometry to gauge selective auditory attention, in conjunction with diverse measures, can significantly enhance our understanding of listeners' requirements (Tessendorf et al., 2011; Zhang et al., 2021). As proposed by Alhanbali et al. (2021), utilizing pupil diameter can vastly improve hearing rehabilitation through devices aimed at addressing hearing impairments. The integration of wearable eye trackers with hearing aids, as seen in the concept of an attentive hearing aid (Tessendorf et al., 2011), exemplifies a transformative application. This setup allows hearing-impaired individuals to selectively filter sound sources based on eye movement, enabling improved communication in various settings. Furthermore, insights from Ledoux et al. (2016) underscore the potential to discern whether a listener comprehends a spoken word by analyzing pupil diameter. This avenue opens doors to understanding language development in individuals with autism spectrum disorder and communication challenges, offering pathways to enhance their communication skills.

Moreover, the exciting prospect of delving into directional selective auditory attention through integration with eye trackers in virtual reality (VR) headsets holds immense potential for future research (Imaoka et al., 2020, Chen et al., 2017). This technology fusion provides an immersive and controlled environment, enabling a deeper understanding of how individuals selectively attend to auditory stimuli in various directions. As we consistently advance and fine-tune these methodologies, an array of possibilities emerges. From refining assistive technologies to developing novel strategies, we can pave the way for enhanced auditory experiences. These advancements are crucial for catering to a wide spectrum of individuals across diverse contexts, ultimately promoting inclusivity and optimizing auditory support for those with unique hearing needs. The fusion of eye tracking technology with VR not only offers research opportunities but also holds promise for real-world applications, revolutionizing how we approach auditory attention and ultimately enriching the lives of individuals with varying hearing abilities.

6.2. Limitations and Future Work

This chapter concludes by outlining avenues for future research along with offering recommendations, while also acknowledging the limitations inherent in this study.

Limitations were embraced as stepping stones to refinement. The constraint of two speakers, though certified narrators, possibly introduced pronunciation biases. The simplified left, right, and bilateral audio directions, though strategic, necessitated further exploration in the context of binaural listening. Despite these constraints, the study resonated with real-life complexity, akin to the challenges posed by multiple sound sources, like the cacophony of a bustling cocktail party.

The distinction that the sounds are delivered separately to the right and left ears, rather than being binaural, presents a noteworthy limitation in our study. This setup could potentially influence how participants perceive and process the auditory stimuli. To mitigate this limitation and enhance the study's design, one suggestion is to replicate the experiment with a true binaural setup, where sounds are presented simultaneously to both ears, mimicking a more natural listening experience.

One limitation of our study is the use of monosyllabic words to assess auditory attention. Monosyllabic words, while useful for their simplicity and efficiency, may not fully represent the complexities of language processing and attention allocation in real-world scenarios. To address this limitation, future research could incorporate more diverse and complex linguistic stimuli, such as polysyllabic words, sentences, or even discourse, to provide a more comprehensive understanding of auditory attention across various linguistic contexts. Additionally, employing ecologically valid auditory stimuli that mimic everyday listening situations could enhance the generalizability of findings and better reflect real-world attention demands.

A potential limitation of this study may lie in the quality and specifications of the headphones used for delivering auditory stimuli. The choice of headphones can significantly influence the fidelity and accuracy of sound delivery, which in turn can affect participants' perception and interpretation of the auditory materials. To address this concern and enhance the study's robustness, it is advisable to employ high-quality headphones with superior sound reproduction capabilities.

Nonetheless, this journey is but a step in a larger odyssey. Future avenues beckon with the promise of expanding the study's scope. The inclusion of more speakers, binaural listening scenarios, and diversified sound sources could further enrich our understanding. The prospect of leveraging virtual reality headsets and integrated eye tracking systems opens exciting possibilities for probing selective auditory attention in immersive environments.

Further studies should explore the correlation between PESQ values, SII values, and pupil dilation across various languages.

The study participants consisted of individuals at the university level, spanning ages 18 to 41. As the analysis focused exclusively on this age range, the insights derived from the results primarily apply to university students and adult participants.

REFERENCES

Ahern, S., & Beatty, J. (1979). Pupillary responses during information processing vary with Scholastic Aptitude Test scores. *Science, 205*(4412), 1289-1292.

Aksan, Y., Aksan, M., Koltuksuz, A., Sezer, T., Mersinli, Ü., Demirhan, U. U., ... & Kurtoğlu, Ö. (2012, May). Construction of the Turkish national corpus (TNC). In Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12) (pp. 3223-3227).

Alhanbali, S., Munro, K. J., Dawes, P., Carolan, P. J., & Millman, R. E. (2021). Dimensions of self-reported listening effort and fatigue on a digits-in-noise task, and association with baseline pupil size and performance accuracy. International Journal of Audiology, 60(10), 762-772.

Ali, F. B., Larbi, S. D., Jaïdane, M., & Ridane, K. (2009, August). Experimental mappings and validation of the dependence on the language of objective speech quality scores in actual GSM network conditions. In *2009 17th European Signal Processing Conference* (pp. 2534-2538). IEEE.

Altmann, G. T., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition, 73*(3), 247-264.

Arias, S., Mossberg, A., Nilsson, D., & Wahlqvist, J. (2021). A Study on Evacuation Behavior in Physical and Virtual Reality Experiments. In *Fire Technology*. Springer US. https://doi.org/10.1007/s10694-021-01172-4

Asbjornsen, A. E., & Hugdahl, K. (1995). Attentional effects in dichotic listening. Brain and language, 49(3), 189-201.

Aston-Jones, G., & Cohen, J. D. (2005). An Integrative Theory Of Locus Coeruleus-Norepinephrine Function: Adaptive Gain and Optimal Performance. *Annual Review of Neuroscience*, *28*(1), 403–450.

Ayasse, N. D., Lash, A., & Wingfield, A. (2017). Effort Not Speed Characterizes Comprehension of Spoken Sentences by Older Adults with Mild Hearing Impairment. *Frontiers in Aging Neuroscience*, *8*, 329. https://doi.org/10.3389/fnagi.2016.00329

Bakry, A., Al-khatib, R., Negm, R., Sabra, E., Maher, M., Mohamed, Z., Shawky, D., & Badawi, A. (2020). Using Eye Movement to Assess Auditory Attention. In A. E. Hassanien, A. T. Azar, T. Gaber, R. Bhatnagar, & M. F. Tolba (Eds.), *The International Conference on Advanced Machine Learning Technologies and Applications (AMLTA2019)* (pp. 200–208). Springer International Publishing.

Balota, D. A., & Chumbley, J. I. (1984). Are lexical decisions a good measure of lexical access? The role of word frequency in the neglected decision stage. Journal of Experimental Psychology: Human perception and performance, 10(3), 340.

Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., ... & Green, P. (2009). Package 'lme4'. *URL http://lme4. r-forge. r-project. org*.

Beatty, J. (1982). Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological bulletin*, *91*(2), 276.

Beatty, J., & Kahneman, D. (1966). Pupillary changes in two memory tasks. *Psychonomic Science*, *5*(10), 371-372.

Beatty, J., & Lucero-Wagoner, B. (2000). The pupillary system.

Beerends, J. G., Hekstra, A. P., Rix, A. W., & Hollier, M. P. (2002). Perceptual evaluation of speech quality (pesq) the new itu standard for end-to-end speech quality assessment part ii: psychoacoustic model. *Journal of the Audio Engineering Society*, *50*(10), 765-778.

Beerends, J. G., Larsen, E., Iyer, N., & Van Vugt, J. (2004, June). Measurement of speech intelligibility based on the PESQ approach. In *Proceedings of the Workshop Measurement of Speech and Audio Quality in Networks (MESAQIN), Prague, Czech Republic*.

Beerends, J. G., van Wijngaarden, S., & van Buuren, R. (2005). Extension of ITU-T recommendation P. 862 PESQ towards measuring speech intelligibility with vocoders. *New Directions for Improving Audio Effectiveness*, 10-1.

Beerends, J., van Buuren, R. A., Van Vugt, J. M., & Verhave, J. A. (2009, March). PESQ based speech intelligibility measurements. In *Proceedings of the International Conference on Acoustics including the German 35th Annual Conference on Acoustics NAG/DAGA. Rotterdam (Holland): TNO*.

Bengio, Y., Ducharme, R., Vincent, P., & Janvin, C. (2003). A Neural Probabilistic Language Model. The Journal of Machine Learning Research, 3, 1137–1155. http://doi.org/10.1162/153244303322533223

Berti, S. (2012). Automatic processing of rare versus novel auditory stimuli reveal different mechanisms of auditory change detection. *NeuroReport,* 23, 441–446. https://doi.org/10.1097/WNR.0b013e32835308b5.

Beutelmann, R., & Brand, T. (2006). Prediction of speech intelligibility in spatial noise and reverberation for normal-hearing and hearing-impaired listeners. *The Journal of the Acoustical Society of America*, *120*(1), 331-342.Print.

Bojko, A. (2013). *Eye tracking the user experience: A practical guide to research*. Rosenfeld Media.

Boothroyd, A., Erickson, F. N., & Medwetsky, L. (1994). The hearing aid input: A phonemic approach to assessing the spectral distribution of speech. *Ear and Hearing*, *15*(6), 432-442.

Braga, R. M., Fu, R. Z., Seemungal, B. M., Wise, R. J., & Leech, R. (2016). Eye movements during auditory attention predict individual differences in dorsal attention network activity. *Frontiers in human neuroscience,* 10, 164.

Brand, T. (2008). Speech Intelligibility. In: Havelock, D., Kuwano, S., Vorländer, M. (eds) Handbook of Signal Processing in Acoustics. Springer, New York, NY. https://doi.org/10.1007/978-0-387-30441-0_13

Breland, H. M. (1996). Word frequency and word difficulty: A comparison of counts in four corpora. *Psychological Science*, *7*(2), 96-99.

Brouwer, S., Mitterer, H., & Huettig, F. (2012). Speech reductions change the dynamics of competition during spoken word recognition. *Language and Cognitive Processes*, *27*(4), 539-571.

Brouwer, S., Mitterer, H., & Huettig, F. (2013). Discourse context and the recognition of reduced and canonical spoken words. *Applied Psycholinguistics*, *34*(3), 519-539.

Brungart, D. S., & Simpson, B. D. (2007). Cocktail party listening in a dynamic multitalker environment. Perception & Psychophysics, 69(1), 79–91. https://doi.org/10.3758/BF03194455

Burley, D. T., Gray, N. S., & Snowden, R. J. (2017). As Far as the Eye Can See: Relationship between Psychopathic Traits and Pupil Response to Affective Stimuli. PLOS ONE, 12(1), e0167436-. https://doi.org/10.1371/journal.pone.0167436

Cahill, M. (2010). Efficacy of the Speech Intelligibility Index as an intervention tool with pediatric cancer patients.

Challis, B. H., & Krane, R. V. (1988). Mood induction and the priming of semantic memory in a lexical decision task: Asymmetric effects of elation and depression. *Bulletin of the Psychonomic Society*, *26*(4), 309-312.

Chaney, R. H., Givens, C. A., Aoki, M. F., & Gombiner, M. L. (1989). Pupillary Responses in Recognizing Awareness in Persons with Profound Mental Retardation. Perceptual and Motor Skills, 69(2), 523–528.

Chapman, L. R., & Hallowell, B. (2015). A novel pupillometric method for indexing word difficulty in individuals with and without aphasia. *Journal of Speech, Language, and Hearing Research, 58*(5), 1508-1520.

Chen, J., Huang, Q., & Wu, X. (2016). Frequency importance function of the speech intelligibility index for Mandarin Chinese. *Speech Communication*, *83*, 94-103.

Chen, H., Dey, A., Billinghurst, M., & Lindeman, R. W. (2017). Exploring pupil dilation in emotional virtual reality environments. *International Conference on Artificial Reality and Telexistence and Eurographics Symposium on Virtual Environments, ICAT-EGVE 2017*, 169–176. https://doi.org/10.2312/egve.20171355

Cherry, E. C. (1953). Some Experiments on the Recognition of Speech, with One and with Two Ears. *The Journal of the Acoustical Society of America*, *25*(5), 975–979. https://doi.org/10.1121/1.1907229

Chong, F. L., McLoughlin, I. V., & Pawlikowski, K. (2005, November). A methodology for improving PESQ accuracy for Chinese speech. In *TENCON 2005-2005 IEEE Region 10 Conference* (pp. 1-6). IEEE.

Cohen S., Evans G.W., Krantz D.S. & Stokols, D. 1980. Physiological, motivational, and cognitive effects of aircraft noise on children. Moving from the laboratory to the fi eld. *Am Psychol* , 35, 3, 231 – 243.

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: a dual route cascaded model of visual word recognition and reading aloud. *Psychological review*, *108*(1), 204.

Côté, N. (2011). *Integral and diagnostic intrusive prediction of speech quality*. Springer Science & Business Media.

Dabbs, J. M. (1997). Testosterone and Pupillary Response to Auditory Sexual Stimuli. *Physiology & Behavior*, 62(4), 909–912.

Damico, J. S., & Ball, M. J. (Eds.). (2019). *The SAGE encyclopedia of human communication sciences and disorders*. SAGE Publications.

Dambacher, M., Kliegl, R., Hofmann, M., & Jacobs, A. M. (2006). Frequency and predictability effects on event-related potentials during reading. *Brain research*, *1084*(1), 89-103.

Drager, K. D., & Reichle, J. E. (2001). Effects of discourse context on the intelligibility of synthesized speech for young adult and older adult listeners: Applications for AAC. *Journal of Speech, Language, and Hearing Research, 44*(5), 1052-1057.

Dupuy, H. J. (1974). The rationale, development, and standardization of a basic word vocabulary test.

Eckert, M., Habets, E. A. P., & Rummukainen, O. S. (2021). Cognitive load estimation based on pupillometry in virtual reality with uncontrolled scene lighting. *2021 13th International Conference on Quality of Multimedia Experience, QoMEX 2021*, 73–76. https://doi.org/10.1109/QoMEX51781.2021.9465417

Elshtain, E. L., & Schaefer, T. (1968). Effects of storage load and word frequency on pupillary responses during short-term memory. *Psychonomic Science, 12*(4), 143-144.

Enders, L. R., Smith, R. J., Gordon, S. M., Ries, A. J., & Touryan, J. (2021). Gaze Behavior During Navigation and Visual Search of an Open-World Virtual Environment. *Frontiers in Psychology*, *12*(August), 1–18. https://doi.org/10.3389/fpsyg.2021.681042

Foundas, A. L., Corey, D. M., Hurley, M. M., & Heilman, K. M. (2006). Verbal dichotic listening in right and left-handed adults: Laterality effects of directed attention. *Cortex*, *42*(1), 79–86. https://doi.org/10.1016/S0010-9452(08)70324-1

Freyman, R. L., Balakrishnan, U., & Helfer, K. S. (2004). Effect of number of masking talkers and auditory priming on informational masking in speech recognition. The Journal of the Acoustical Society of America, 115(5), 2246–2256. https://doi.org/10.1121/1.1689343

Gagné, J. P., Besser, J., & Lemke, U. (2017). Behavioral Assessment of Listening Effort Using a Dual-Task Paradigm. *Trends in hearing*, *21*, 2331216516687287. https://doi.org/10.1177/2331216516687287

Galbrun, L., & Kitapci, K. (2016). Speech intelligibility of English, Polish, Arabic and Mandarin under different room acoustic conditions. *Applied Acoustics*, *114*, 79-91.

Gatehouse, S., & Noble, W. (2004). The speech, spatial and qualities of hearing scale (SSQ). *International journal of audiology*, *43*(2), 85-99

Geal-Dor,, M., Kamenir,, Y., & Babkoff,, H. (2005). Event Related Potentials (ERPs) and Behavioral Responses: Comparison of Tonal Stimuli to Speech Stimuli in Phonological and Semantic Tasks. *Journal of Basic and Clinical Physiology and Pharmacology*, 16, 139 - 156. https://doi.org/10.1515/JBCPP.2005.16.2-3.139.

Gernsbacher, M. A. (1984). Resolving 20 years of inconsistent interactions between lexical familiarity and orthography, concreteness, and polysemy. *Journal of experimental psychology: General*, *113*(2), 256.

Ghosh, P., Mazumder, A., Bhattacharyya, S., & Tibarewala, D. N. (2015). An EEG study on working memory and cognition. *ACM International Conference Proceeding Series*, *26-27-February-2015*, 21–26. https://doi.org/10.1145/2708463.2709065

Gilzenrat, M. S., Nieuwenhuis, S., Jepma, M., & Cohen, J. D. (2010). Pupil diameter tracks changes in control state predicted by the adaptive gain theory of locus coeruleus

function. *Cognitive, Affective, & Behavioral Neuroscience*, *10*(2), 252–269. https://doi.org/10.3758/CABN.10.2.252

Gingras, B., Marin, M. M., Puig-Waldmüller, E., & Fitch, W. T. (2015). The Eye is Listening: Music-Induced Arousal and Individual Differences Predict Pupillary Responses. Frontiers in Human Neuroscience, 9, 619. https://doi.org/10.3389/fnhum.2015.00619

Gopher, D. (1973). Eye-movement patterns in selective listening tasks of focused attention. Perception & Psychophysics, 14(2), 259–264.

Grange, J. A., & Culling, J. F. (2016). The benefit of head orientation to speech intelligibility in noise. *The Journal of the Acoustical Society of America*, *139*(2), 703–712. https://doi.org/10.1121/1.4941655

Granholm, E., Asarnow, R. F., Sarkin, A. J., & Dykes, K. L. (1996). Pupillary responses index cognitive resource limitations. *Psychophysiology*, *33*(4), 457–461. https://doi.org/https://doi.org/10.1111/j.1469-8986.1996.tb01071.x

Guéguin, M., Le Bouquin-Jeannès, R., Gautier-Turbin, V., Faucon, G., & Barriac, V. (2008). On the evaluation of the conversational speech quality in telecommunications. *EURASIP Journal on Advances in Signal Processing*, *2008*, 1-15.

Hällgren, M., Larsby, B., Lyxell, B., & Arlinger, S. (2005). Speech understanding in quiet and noise, with and without hearing aids: Comprensión del lenguaje en silencio y con ruido, con y sin auxiliares auditivos. *International Journal of Audiology*, *44*(10), 574-583.

Hauk, O., & Pulvermüller, F. (2004). Effects of word length and frequency on the human event-related potential. *Clinical Neurophysiology*, *115*(5), 1090-1103.

Hendrikse, M. M. E., Llorach, G., Grimm, G., & Hohmann, V. (2018). Influence of visual cues on head and eye movements during listening tasks in multi-talker audiovisual environments with animated characters. *Speech Communication*, *101*, 70–84. https://doi.org/https://doi.org/10.1016/j.specom.2018.05.008

Hess, E. H., & Polt, J. M. (1960). Pupil size as related to interest value of visual stimuli. *Science, 132*(3423), 349-350.

Hu, Y., & Loizou, P. C. (2007). Evaluation of objective quality measures for speech enhancement. *IEEE Transactions on audio, speech, and language processing*, *16*(1), 229-238.

Huettig, F., Rommers, J., & Meyer, A. S. (2011). Using the visual world paradigm to study language processing: A review and critical evaluation. *Acta psychologica, 137*(2), 151-171.

Hyönä, J., Tommola, J., & Alaja, A.-M. (1995). Pupil Dilation as a Measure of Processing Load in Simultaneous Interpretation and Other Language Tasks. *The Quarterly Journal of Experimental Psychology Section A*, *48*(3), 598–612. https://doi.org/10.1080/14640749508401407

IEEE Trans. Audio Speech Lang Processing 16(1), 229–238 (2008)

Imaoka, Y., Flury, A., & de Bruin, E. D. (2020). Assessing Saccadic Eye Movements With Head-Mounted Display Virtual Reality Technology. *Frontiers in Psychiatry*, *11*, 0–19. https://doi.org/10.3389/fpsyt.2020.572938

ITU-T P.800, 1996, ITU-T Recommendation P.800. 1996. Methods for subjective determination of transmission quality (formerly Rec. P.80)

ITU-T Rec. P.861, Recommendation, P. (1998). 861: Objective Quality Measurement of Telephone-Band (300-3400 Hz) Speech Codecs. *International Telecommunications Union-Telecommunication Standardization Sector Std*.

ITU-T Recommendation P.563, 2004, Union, I. T. (2004). Single ended method for objective speech quality assessment in narrow-band telephony applications. *ITU-T Recommendation*, 563.

Jin, I. K., Kates, J. M., & Arehart, K. H. (2017). Sensitivity of the Speech Intelligibility Index to the assumed dynamic range. *Journal of Speech, Language, and Hearing Research*, *60*(6), 1674-1680.

Jin, I. K. (2014). *Development of the speech intelligibility index (SII) for Korean* (Doctoral dissertation, University of Colorado at Boulder)

Jin, I. K., Kates, J. M., & Arehart, K. H. (2017). Does language matter when using a graphical method for calculating the Speech Intelligibility Index?. *Journal of the American Academy of Audiology*, *28*(02), 119-126.

Jin, P., Zou, J., Zhou, T., & Ding, N. (2018). Eye activity tracks task-relevant structures during speech and auditory sequence perception. *Nature Communications*, *9*(1), 5374. https://doi.org/10.1038/s41467-018-07773-y

Kahneman, D., & Beatty, J. (1966). Pupil Diameter and Load on Memory. *Science*, *154*(3756), 1583–1585. https://doi.org/10.1126/science.154.3756.1583

Kalikow, D. N., Stevens, K. N. and Elliott, L. L. 1977. Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. *Journal of the Acoustical Society of America*, 61: 1337–1351.

Kawahara, T., Inoue, K., Lala, D., & Takanashi, K. (2018). Audio-Visual Conversation Analysis by Smart Posterboard and Humanoid Robot. *2018 IEEE*

International Conference on Acoustics, Speech and Signal Processing (ICASSP), 6573–6577. https://doi.org/10.1109/ICASSP.2018.8461470

Kent, R. D., Weismer, G., & Kent, J. F. (1989). Toward phonetic intelligibility testing in dysarthria. Journal of Speech and Hearing Disorders, 54, 482–499.

Kelime Listelerinde Arama. (2023, August 28). Retrieved from Kelimeler.Net: https://kelimeler.net/

Kidd, G., Arbogast, T. L., Mason, C. R., & Gallun, F. J. (2005). The advantage of knowing where to listen. *The Journal of the Acoustical Society of America*, *118*(6), 3804–3815. https://doi.org/10.1121/1.2109187

Kim, C., Mason, R., & Brookes, T. (2013). Head movements made by listeners in experimental and real-life listening activities. Journal of the Audio Engineering Society, 61(6 (Jun), 425–438.

Kirişoğlu, Ş. (2013, May 21). Türkçe tek heceli kelimeler (işitme ve konuşma terapisi boyama çalışmaları). Retrieved from Ody ve Kon Boz Uzm. ŞULE TÜFEKÇİOĞLU: http://sulekisioglu.blogspot.com/2013/05/tek-heceli-kelimeler.html

Kitterick, P. T., Bailey, P. J., & Summerfield, A. Q. (2010). Benefits of knowing who, where, and when in multi-talker listening. *The Journal of the Acoustical Society of America*, *127*(4), 2498–2508. https://doi.org/10.1121/1.3327507

Koch, I., Lawo, V., Fels, J., & Vorländer, M. (2011). Switching in the Cocktail Party: Exploring Intentional Control of Auditory Selective Attention. *Journal of Experimental Psychology: Human Perception and Performance*, *37*(4), 1140–1147. https://doi.org/10.1037/a0022189

Koch, X., & Janse, E. (2016). Speech rate effects on the processing of conversational speech across the adult life span. *The Journal of the Acoustical Society of America*, *139*(4), 1618–1636. https://doi.org/10.1121/1.4944032

Koelewijn, T., Shinn-Cunningham, B.G., Zekveld, A.A., Kramer, S.E., (2014). The pupil response is sensitive to divided attention during speech processing. Hear. Res.312, 114e120.

Koelewijn, T., de Kluiver, H., Shinn-Cunningham, B. G., Zekveld, A. A., & Kramer, S. E. (2015). The pupil response reveals increased listening effort when it is difficult to focus attention. *Hearing Research*, *323*, 81–90.

Koelewijn, T., Shinn-Cunningham, B. G., Zekveld, A. A., & Kramer, S. E. (2014). The pupil response is sensitive to divided attention during speech processing. *Hearing Research*, *312*, 114–120. https://doi.org/https://doi.org/10.1016/j.heares.2014.03.010

Koelewijn, T., Versfeld, N. J., & Kramer, S. E. (2017). Effects of attention on the speech reception threshold and pupil response of people with impaired and normal hearing. *Hearing Research*, *354*, 56–63. https://doi.org/10.1016/J.HEARES.2017.08.006

Koelewijn, T., Zekveld, A. A., Festen, J. M., & Kramer, S. E. (2012). Pupil Dilation Uncovers Extra Listening Effort in the Presence of a Single-Talker Masker. *Ear and Hearing*, *33*(2).

Kompus, K., Specht, K., Ersland, L., Juvodden, H. T., van Wageningen, H., Hugdahl, K., & Westerhausen, R. (2012). A forced-attention dichotic listening fMRI study on 113 subjects. *Brain and Language*, *121*(3), 240–247. https://doi.org/10.1016/j.bandl.2012.03.004

Kramer, S. E., Lorens, A., Coninx, F., Zekveld, A. A., Piotrowska, A., & Skarzynski, H. (2013). Processing load during listening: The influence of task characteristics on the pupil response. *Language and cognitive processes*, *28*(4), 426-442.

Kuchinke, L., Võ, M. L. H., Hofmann, M., & Jacobs, A. M. (2007). Pupillary responses during lexical decisions vary with word frequency but not emotional valence. *International Journal of Psychophysiology*, *65*(2), 132-140.

Kuchinke, L., Jacobs, A. M., Grubich, C., Vo, M. L. H., Conrad, M., & Herrmann, M. (2005). Incidental effects.

Kuchinsky, S. E., Ahlstrom, J. B., Vaden Jr, K. I., Cute, S. L., Humes, L. E., Dubno, J. R., & Eckert, M. A. (2013). Pupil size varies with word listening and response selection difficulty in older adults with hearing loss. *Psychophysiology*, *50*(1), 23-34.

Kuchinsky, S. E., Vaden, K. I., Ahlstrom, J. B., Cute, S. L., Humes, L. E., Dubno, J. R., & Eckert, M. A. (2016). Task-Related Vigilance During Word Recognition in Noise for Older Adults with Hearing Loss. *Experimental Aging Research*, *42*(1), 50–66. https://doi.org/10.1080/0361073X.2016.1108712

Kuipers, J. R., & Thierry, G. (2011). N400 amplitude reduction correlates with an increase in pupil size. *Frontiers in Human Neuroscience*, *5*, 61.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. (2017). lmerTest package: tests in linear mixed effects models. *Journal of statistical software*, *82*, 1-26.

Kwok, C. K. (2018). *During Tasks Via Facial Responses, Eye Gaze And Mouse Movements.* https://theses.lib.polyu.edu.hk/handle/200/9304

Köster, F., Schiffner, F., Möller, S., & Malfait, L. (2017). Towards degradation decomposition for voice communication system assessment. *Quality and User Experience*, *2*(1), 1-22.

Laeng, B., Sirois, S., & Gredebäck, G. (2012). Pupillometry: A Window to the Preconscious? *Perspectives on Psychological Science*, *7*(1), 18–27. https://doi.org/10.1177/1745691611427305

Larsby, B., Ha¨llgren, M., Lyxell, B. & Arlinger, S. 2005. Cognitive performance and perceived effort in speech processing tasks: effects of different noise backgrounds in normals and in hearing-impaired subjects. Int J Audiol, 44, 131/143.

Landgraf, S., Van der Meer, E., & Krueger, F. (2010). Cognitive resource allocation for neural activity underlying mathematical cognition: a multi-method study. *Zdm, 42*(6), 579-590.

Ledoux, K., Coderre, E., Bosley, L., Buz, E., Gangopadhyay, I., & Gordon, B. (2016). The concurrent use of three implicit measures (eye movements, pupillometry, and event-related potentials) to assess receptive vocabulary knowledge in normal adults. *Behavior Research Methods*, *48*(1), 285–305. https://doi.org/10.3758/s13428-015-0571-6

Lee, S. P., Badler, J. B., & Badler, N. I. (2002). Eyes Alive. *Proceedings of the 29th Annual Conference on Computer Graphics and Interactive Techniques*, 637–644. https://doi.org/10.1145/566570.566629

Lee, S., Mendel, L. L., & Bidelman, G. M. (2019). Predicting speech recognition using the speech intelligibility index and other variables for cochlear implant users. *Journal of Speech, Language, and Hearing Research, 62*(5), 1517-1531.

Leigh, R. J., & Zee, D. S. (1991). *The neurology of eye movements*. F.A. Davis Co.

Letowski, T. R., & Scharine, A. A. (2017). *Correlational analysis of speech intelligibility tests and metrics for speech transmission*. US Army Research Laboratory Aberdeen Proving Ground United States.

Litovsky Ruth Y.and Goupell, M. J. and M. S. M. and K. A. (2017). Hearing with Cochlear Implants and Hearing Aids in Complex Auditory Scenes. In J. Z. and P. A. N. and F. R. R. Middlebrooks John C.and Simon (Ed.), *The Auditory System at the Cocktail Party* (pp. 261–291). Springer International Publishing. https://doi.org/10.1007/978-3-319-51662-2_10

Liu, W. M., Jellyman, K. A., Mason, J. S., & Evans, N. W. (2006, May). Assessment of objective quality measures for speech intelligibility estimation. In *2006 IEEE International Conference on Acoustics Speech and Signal Processing Proceedings* (Vol. 1, pp. I-I). IEEE.

Liu, W. M. (2008). *Objective assessment of speech intelligibility*. Swansea University (United Kingdom).

Loizou, P. C. (2007). *Speech enhancement: theory and practice*. CRC press.

Loizou, P. C. (2011). Speech quality assessment. In *Multimedia analysis, processing and communications* (pp. 623-654). Springer, Berlin, Heidelberg.

Ma, J., Hu, Y., & Loizou, P. C. (2009). Objective measures for predicting speech intelligibility in noisy conditions based on new band-importance functions. *The Journal of the Acoustical Society of America*, *125*(5), 3387-3405.

Mackersie C.L. & Cones H. 2011. Subjective and psychophysiological indices of listening effort in a competing-talker task. J Am Acad Audiol , 22 , 113 – 122 .

Mathôt, S., Fabius, J., Van Heusden, E., & Van der Stigchel, S. (2018). Safe and sensible preprocessing and baseline correction of pupil-size data. *Behavior Research Methods*, *50*(1), 94-106.

Mattys, S. L., Davis, M. H., Bradlow, A. R., & Scott, S. K. (2012). Speech recognition in adverse conditions: A review. *Language and cognitive processes*, *27*(7-8), 953-978.

McCloy, D. R., Lau, B. K., Larson, E., Pratt, K. A. I., & Lee, A. K. C. (2017). Pupillometry shows the effort of auditory attention switching. *The Journal of the Acoustical Society of America*, *141*(4), 2440–2451. https://doi.org/10.1121/1.4979340

McGarrigle, R., Dawes, P., Stewart, A. J., Kuchinsky, S. E., & Munro, K. J. (2017). Measuring listening-related effort and fatigue in school-aged children using pupillometry. *Journal of Experimental Child Psychology*, *161*, 95-112.

Michie, P., Bearpark, H., Crawford, J., & Glue, L. (1990). The nature of selective attention effects on auditory event-related potentials. *Biological Psychology*, 30, 219-250. https://doi.org/10.1016/0301-0511(90)90141-I.

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.

Miller, G. A., Heise, G. A., & Lichten, W. (1951). The intelligibility of speech as a function of the context of the test materials. *Journal of experimental psychology*, *41*(5), 329.

Monsell, S., Doyle, M. C., & Haggard, P. N. (1989). Effects of frequency on visual word recognition tasks: Where are they?. *Journal of Experimental Psychology: General*, *118*(1), 43.

Morris, S. K., Granholm, E., Sarkin, A. J., & Jeste, D. V. (1997). Effects of schizophrenia and aging on pupillographic measures of working memory. *Schizophrenia Research*, *27*(2), 119–128.

Nathwani, K., Richard, G., David, B., Prablanc, P., & Roussarie, V. (2017). Speech intelligibility improvement in car noise environment by voice transformation. *Speech Communication*, *91*, 17-27.

Nielsen, J., & Pernice, K. (2010). *Eyetracking Web Usability*. Berkeley, California USA: New Riders.

Ohlenforst, B., Zekveld, A. A., Lunner, T., Wendt, D., Naylor, G., Wang, Y., . . . Kramer, S. E. (2017). Impact of stimulus-related factors and hearing impairment on listening effort as indicated by pupil dilation. *Hearing Research, 351*, 68-79.

Pals, C., Sarampalis, A., van Dijk, M., & Baskent, D. (2019). Effects of additional low-pass–filtered speech on listening effort for noise-band–vocoded speech in quiet and in noise. *Ear and hearing*, *40*(1), 3-17.

Papesh, M. H., Goldinger, S. D., & Hout, M. C. (2012). Memory strength and specificity revealed by pupillometry. *International Journal of Psychophysiology, 83*(1), 56-64.

Pannese, A., Herrmann, C., & Sussman, E. (2015). Analyzing the Auditory Scene: Neurophysiologic Evidence of a Dissociation Between Detection of Regularity and Detection of Change. *Brain Topography*, 28, 411-422.

Partala, T., & Surakka, V. (2003). Pupil size variation as an indication of affective processing. *International journal of human-computer studies*, *59*(1-2), 185-198.

Pashler, H. E. (1998). *The psychology of attention*. MIT Press.

Pavlov, Y. G., & Kotchoubey, B. (2017). EEG correlates of working memory performance in females. *BMC Neuroscience*, *18*(1), 26.

Pearsons, K. S., Bennett, R. L., & Fidell, S. A. (1977). *Speech levels in various noise environments*. Office of Health and Ecological Effects, Office of Research and Development, US EPA.

Peelle, J. E. (2018). Listening Effort: How the Cognitive Consequences of Acoustic Challenge Are Reflected in Brain and Behavior. *Ear and Hearing*, *39*(2), 204–214. https://doi.org/10.1097/AUD.0000000000000494

Pichora-Fuller, M. K., Kramer, S. E., Eckert, M. A., Edwards, B., Hornsby, B. W., Humes, L. E., . . . Mackersie, C. L. (2016). Hearing impairment and cognitive energy: The framework for understanding effortful listening (FUEL). *Ear and Hearing, 37*, 5S-27S.

Piquado, T., Isaacowitz, D., & Wingfield, A. (2010). Pupillometry as a measure of cognitive effort in younger and older adults. *Psychophysiology*, *47*(3), 560-569.

Plomp, R. (1986). A signal-to-noise ratio model for the speech-reception threshold of the hearing impaired. *Journal of Speech, Language, and Hearing Research*, *29*(2), 146-154.

Poole, A., & Ball, L. J. (2005). Eye Tracking in Human-Computer Interaction and Usability Research: Current Status and Future Prospects. In C. Ghaoui (Ed.). Pennsylvania: Idea Group, Inc.

Putze, F., Küster, D., Annerer-Walcher, S., & Benedek, M. (2018). Dozing Off or Thinking Hard? Classifying Multi-Dimensional Attentional States in the Classroom from Video. *Proceedings of the 20th ACM International Conference on Multimodal Interaction*, 258–262. https://doi.org/10.1145/3242969.3243000

Rayner, K., & Duffy, Sbah. A. (1986). Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. *Memory & cognition*, *14*(3), 191-201.

Rec, I. T. U. T. (2004). P. 563: Single-ended method for objective speech quality assessment in narrow-band telephony applications. *International Telecommunication Union, Geneva*, 1-25.

Rec, I. T. U. T. (1998). P. 861: Objective quality measurement of telephone-band (300–3400 Hz) speech codecs. *International Telecommunication Union, Geneva, Switzerland*.

Reel, L. A., & Hicks, C. B. (2012). Selective Auditory Attention in Adults: Effects of Rhythmic Structure of the Competing Language. *Journal of Speech, Language, and Hearing Research, 55*(1), 89–104. https://doi.org/10.1044/1092-4388(2011/10-0193)

Reilly, J., Kelly, A., Kim, S. H., Jett, S., & Zuckerman, B. (2018). The human task-evoked pupillary response function is linear: Implications for baseline response scaling in pupillometry. *Behavior Research Methods*, 1-14.

Rennies, J., Schepker, H., Holube, I., & Kollmeier, B. (2014). Listening effort and speech intelligibility in listening situations affected by noise and reverberation. *The Journal of the Acoustical Society of America*, *136*(5), 2642-2653.

Rix, A. W., Beerends, J. G., Hollier, M. P., & Hekstra, A. P. (2001, May). Perceptual evaluation of speech quality (PESQ)-a new method for speech quality assessment of telephone networks and codecs. In *2001 IEEE international conference on acoustics, speech, and signal processing. Proceedings (Cat. No. 01CH37221)* (Vol. 2, pp. 749-752). IEEE.

Rhebergen, K. S., & Versfeld, N. J. (2005). A speech intelligibility index-based approach to predict the speech reception threshold for sentences in fluctuating noise for normal-hearing listeners. *The Journal of the Acoustical Society of America*, *117*(4), 2181-2192.

Rolph Houben, Maaike van Doorn-Bierman & Wouter A. Dreschler (2013). Using response time to speech as a measure for listening effort, International Journal of Audiology, 52:11, 753-761, DOI: 10.3109/14992027.2013.832415

Rugg, M. D. (1990). Event-related brain potentials dissociate repetition effects of high- and low-frequency words. *Memory & cognition*, *18*(4), 367-379.

Sağlam, F., Genc, B., & Sever, H. (2019). Extending a sentiment lexicon with synonym--antonym datasets: SWNetTR++. *Turkish Journal of Electrical Engineering and Computer Sciences*, *27*(3), 1806-1820.

Sahlén, B., Haake, M., von Lochow, H., Holm, L., Kastberg, T., Brännström, K. J., & Lyberg-Åhlander, V. (2018). Is children's listening effort in background noise influenced by the speaker's voice quality?. *Logopedics Phoniatrics Vocology*, *43*(2), 47-55.

Salvaggio, S., & Masson, N. (2019). Eye Position Reflects the Spatial Coding of Numbers During Magnitude Comparison. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. https://doi.org/10.1037/xlm0000681.supp

Schepker, H., Haeder, K., Rennies, J., & Holube, I. (2016). Perceived listening effort and speech intelligibility in reverberation and noise for hearing-impaired listeners. *International journal of audiology*, *55*(12), 738-747.

Schmidtke, J. (2014). Second language experience modulates word retrieval effort in bilinguals: Evidence from pupillometry. *Frontiers in psychology*, *5*, 137.

Schröger, E. (1996). A neural mechanism for involuntary attention shifts to changes in auditory stimulation.. *Journal of cognitive neuroscience*, 8 6, 527-39 . https://doi.org/10.1162/jocn.1996.8.6.527.

Scollie, S. D. (2008). Children's speech recognition scores: The Speech Intelligibility Index and proficiency factors for age and hearing level. *Ear and hearing*, *29*(4), 543-556.

Schluroff, M. (1982). Pupil responses to grammatical complexity of sentences. *Brain and Language, 17*(1), 133-145.

Sereno, S. C., Rayner, K., & Posner, M. I. (1998). Establishing a time-line of word recognition: evidence from eye movements and event-related potentials. *Neuroreport*, *9*(10), 2195-2200.

Viswanathan, D., Wang, Y., Naylor, P. A., & Brookes, M. (2016). A data-driven non-intrusive measure of speech quality and intelligibility. *Speech Communication, 80*, 84-94.

Shimizu, Y., Ohnishi, A., Terada, T., & Tsukamoto, M. (2021). Gaze-Adaptive Subtitles Considering the Balance among Vertical/Horizontal and Depth of Eye Movement. *2021 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, 127–132. https://doi.org/10.1109/ismar-adjunct54149.2021.00035

Siegle, G. J., Granholm, E., Ingram, R. E., & Matt, G. E. (2001). Pupillary and reaction time measures of sustained processing of negative information in depression. *Biological psychiatry*, *49*(7), 624-636.

Siegle, G. J., Ingram, R. E., & Matt, G. E. (2002). Affective interference: An explanation for negative attention biases in dysphoria? *Cognitive Therapy and Research*, *26*(1), 73-87.

Smeds, K., Wolters, F., & Rung, M. (2015). Estimation of signal-to-noise ratios in realistic sound scenarios. *Journal of the American Academy of Audiology*, *26*(02), 183-196.

Siyaev, A., & Jo, G. S. (2021). Towards aircraft maintenance metaverse using speech interactions with virtual objects in mixed reality. *Sensors*, *21*(6), 1–21. https://doi.org/10.3390/s21062066

SR Research Experiment Builder 2.2.1 [Computer software]. (2019). http://www.sr-research.com/

SR Research EyeLink® 1000 User Manual 1.5.0. (2009). http://www.sr-research.com/

Tessendorf, B., Bulling, A., Roggen, D., Stiefmeier, T., Feilner, M., Derleth, P., & Tröster, G. (2011). Recognition of Hearing Needs from Body and Eye Movements to Improve Hearing Instruments. In K. Lyons, J. Hightower, & E. M. Huang (Eds.), *Pervasive Computing* (pp. 314–331). Springer Berlin Heidelberg.

Triyason, T., & Kanthamanon, P. (2012, December). Perceptual evaluation of speech quality measurement on Speex codec VoIP with Tonal language Thai. In *International Conference on Advances in Information Technology* (pp. 181-190). Springer, Berlin, Heidelberg.

Van Engen, K. J., & Peelle, J. E. (2014). Listening effort and accented speech. *Frontiers in human neuroscience*, *8*.

Van Engen, K. J., & McLaughlin, D. J. (2018). Eyes and ears: Using eye tracking and pupillometry to understand challenges to speech recognition. *Hearing Research*, *369*, 56–66. https://doi.org/https://doi.org/10.1016/j.heares.2018.04.013

Verney, S. P., Granholm, E., & Dionisio, D. P. (2001). Pupillary responses and processing resources on the visual backward masking task. *Psychophysiology*, *38*(01), 76-83.

Vertegaal, R., Slagter, R., van der Veer, G., & Nijholt, A. (2001). Eye Gaze Patterns in Conversations: There is More to Conversational Agents than Meets the Eyes. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 301–308. https://doi.org/10.1145/365024.365119

Viswanathan, M., & Viswanathan, M. (2005). Measuring speech quality for text-to-speech systems: development and assessment of a modified mean opinion score (MOS) scale. *Computer speech & language*, *19*(1), 55-83.

Voiers, W. D. & Hehmsoth, C. J. (1973). *Research on diagnostic evaluation of speech intelligibility: Speaker Effects on Intelligibility Test Results.* (No. TRACOR-T73-AU-9010-U). TRACOR INC AUSTIN TX.

Walden, B. E., Surr, R. K., Grant, K. W., et al. (2005). Effect of signal-to noise ratio on directional microphone benefit and preference. *J Am Acad Audiol,* 16, 662–676.

Wendt, D., Dau, T., & Hjortkjær, J. (2016). Impact of background noise and sentence complexity on processing demands during sentence comprehension. *Frontiers in psychology*, *7*, 345.

Westerhausen, R., & Kompus, K. (2018). How to get a left-ear advantage: A technical review of assessing brain asymmetry with dichotic listening. In *Scandinavian Journal of Psychology* (Vol. 59, Issue 1, pp. 66–73). Blackwell Publishing Ltd. https://doi.org/10.1111/sjop.12408

Wingfield, A., & Tun, P. A. (2007). Cognitive supports and cognitive constraints on comprehension of spoken language. *Journal of the American Academy of Audiology, 18*(7), 548-558.

Winn, M. B., Wendt, D., Koelewijn, T., & Kuchinsky, S. E. (2018). Best practices and advice for using pupillometry to measure listening effort: An introduction for those who want to get started. *Trends in hearing, 22*, 2331216518800869

Wong, L. L., Ho, A. H., Chua, E. W., & Soli, S. D. (2007). Development of the Cantonese speech intelligibility index. *The Journal of the Acoustical Society of America*, *121*(4), 2350-2361.

Wu, Y. H., & Bentler, R. A. (2010). Impact of visual cues on directional benefit and preference: Part I—Laboratory tests. *Ear and hearing*, *31*(1), 22-34.

Yost, W. A., Dye, R. H., & Sheft, S. (1996). A simulated "cocktail party" with up to three sound sources. *Perception & Psychophysics*, *58*(7), 1026–1036. https://doi.org/10.3758/BF03206830

Zekveld, A. A., Kramer, S. E., & Festen, J. M. (2010). Pupil response as an indication of effortful listening: The influence of sentence intelligibility. *Ear and Hearing, 31*(4), 480-490.

Zekveld, A. A., Rudner, M., Johnsrude, I. S., Festen, J. M., Van Beek, J. H., & Rönnberg, J. (2011). The influence of semantically related and unrelated text cues on the intelligibility of sentences in noise. *Ear and Hearing*, *32*(6), e16-e25.

Zekveld, A. A., Kramer, S. E., & Festen, J. M. (2011). Cognitive Load During Speech Perception in Noise: The Influence of Age, Hearing Loss, and Cognition on the Pupil Response. *Ear and Hearing*, *32*(4).

Zekveld, A. A., & Kramer, S. E. (2014). Cognitive processing load across a wide range of listening conditions: insights from pupillometry. *Psychophysiology, 51*(3), 277-284.

Zekveld, A. A., Rudner, M., Kramer, S. E., Lyzenga, J., & Rönnberg, J. (2014). Cognitive processing load during listening is reduced more by decreasing voice similarity than by increasing spatial separation between target and masker speech. *Frontiers in Neuroscience*, *8*, 88. https://doi.org/10.3389/fnins.2014.00088

Zekveld, A. A., Koelewijn, T., & Kramer, S. E. (2018). The Pupil Dilation Response to Auditory Stimuli: Current State of Knowledge. *Trends in Hearing*, *22*, 2331216518777174. https://doi.org/10.1177/2331216518777174

Zhang, Y., Lehmann, A., & Deroche, M. (2021). Disentangling listening effort and memory load beyond behavioural evidence: Pupillary response to listening effort during a concurrent memory task. *PloS one*, *16*(3), e0233251.

Zheng, L. J., Mountstephens, J., & Teo, J. (2020). Four-class emotion classification in virtual reality using pupillometry. *Journal of Big Data*, *7*(1). https://doi.org/10.1186/s40537-020-00322-9

APPENDICES

7.1. APPENDIX A

Table A.1: Derived Values of SII, SNR, PESQ

List ID	**Word Order**	**SII(Target)**	**SII(Competing)**	**PESQ(Target)**	**PESQ(Competing)**	**SNR(Target)**	**SNR(Competing)**
1	1	0.4093	0.4492	2.545	2.949	-5.2305	10.7807
1	2	0.5674	0.3292	2.018	2.313	1.0502	3.1053
1	3	0.56041	0.3477	2.853	2.011	-0.82038	1.4254
1	4	0.40174	0.48443	2.864	2.170	-0.29869	10.5127
1	5	0.55279	0.31429	2.993	2.288	-3.4449	5.7726
1	6	0.49679	0.37369	2.591	2.518	5.4183	-3.8918
1	7	0.48966	0.39897	2.336	2.217	4.4159	-1.4038
1	8	0.40498	0.46015	2.478	1.447	0.20065	4.4556
1	9	0.43371	0.47301	2.364	2.156	-1.0601	4.5972
1	10	0.26296	0.64045	1.608	3.040	-4.4739	9.0215
2	1	0.17778	0.73891	1.377	3.046	-5.2815	12.5223
2	2	0.63872	0.27673	2.795	2.045	-2.8115	6.4944
2	3	0.44488	0.44675	2.850	2.992	-1.0221	5.1411
2	4	0.53463	0.34077	3.067	1.728	-2.0367	12.7916
2	5	0.44618	0.41764	2.110	2.430	2.8026	-0.37434
2	6	0.55838	0.34563	2.687	1.570	4.7325	-2.1833
2	7	0.5757	0.33828	2.376	2.049	-4.6437	14.4062
2	8	0.33661	0.52852	2.083	2.790	1.4595	2.8582
2	9	0.61288	0.28324	2.891	1.354	-2.0849	3.7326
2	10	0.3153	0.61498	1.732	2.994	0.74014	5.9889
3	1	0.57672	0.30949	1.707	2.459	-1.8852	5.9368
3	2	0.27717	0.60453	1.653	2.919	-0.38561	8.3537
3	3	0.53624	0.31182	1.124	2.492	-4.3408	11.6603
3	4	0.5978	0.2857	2.683	1.878	4.2112	-1.4105
3	5	0.46276	0.445	2.315	2.272	-0.73149	2.6763
3	6	0.33498	0.57281	1.867	2.699	-4.1878	7.4133
3	7	0.39088	0.48655	2.322	2.803	-2.1972	7.3169
3	8	0.35807	0.51995	0.814	2.399	-3.1146	12.2578
3	9	0.41833	0.45145	1.883	1.727	1.6508	1.0332
3	10	0.48009	0.40137	2.952	3.086	3.9164	0.63085
13	1	0.30053	0.6373	1.901	2.502	-1.1664	9.012

List ID	**Word Order**	**SII(Target)**	**SII(Competing)**	**PESQ(Target)**	**PESQ(Competing)**	**SNR(Target)**	**SNR(Competing)**
13	2	0.50479	0.34497	0.848	2.602	-4.1428	11.1013
13	3	0.60517	0.29224	2.871	1.486	-0.44012	8.535
13	4	0.4334	0.48919	2.160	2.055	-5.1286	14.6346
13	5	0.59872	0.30417	2.522	1.623	-0.05392	4.3852
13	6	0.45053	0.45108	2.317	2.626	-3.0915	8.8012
13	7	0.55141	0.3475	1.417	2.261	-2.6877	7.0798
13	8	0.33511	0.54621	2.191	3.008	-6.3423	16.2197
13	9	0.43251	0.42127	2.124	2.696	-4.0622	7.6556
13	10	0.29676	0.62417	2.175	2.806	-4.6669	8.2957
14	1	0.41905	0.43743	2.297	2.452	2.4704	1.6811
14	2	0.33114	0.50921	1.932	1.849	4.8834	1.1534
14	3	0.22657	0.62931	2.511	3.370	-4.5535	11.5976
14	4	0.30463	0.62338	2.394	2.540	-3.7664	5.8525
14	5	0.25927	0.67279	0.949	2.658	-4.4002	9.0795
14	6	0.39965	0.49914	2.781	1.640	-1.5861	9.8107
14	7	0.64603	0.23841	2.939	2.358	-4.4321	7.9815
14	8	0.33741	0.57056	2.633	2.828	0.33349	3.798
14	9	0.3292	0.57736	1.505	2.663	-3.5315	9.3759
14	10	0.26909	0.6295	1.716	2.857	-1.062	7.1898
15	1	0.50039	0.381	3.147	2.423	2.6049	-1.7344
15	2	0.35239	0.54817	2.355	2.707	-2.5058	6.7552
15	3	0.4661	0.39858	2.315	2.115	-2.9577	5.6559
15	4	0.34115	0.55785	1.796	2.215	-0.21596	4.3672
15	5	0.60891	0.31808	2.557	1.436	-3.3709	6.012
15	6	0.45382	0.45977	2.483	1.739	-0.35613	4.0527
15	7	0.70295	0.20916	2.321	2.338	-2.1778	7.6737
15	8	0.245	0.66686	1.697	2.563	-1.8049	7.3414
15	9	0.54953	0.32137	2.629	2.203	2.4578	1.4778
15	10	0.31995	0.57078	2.010	2.956	-3.2372	5.7777
25	1	0.28189	0.61728	2.351	2.911	-4.6134	9.9565
25	2	0.33633	0.58244	1.784	2.808	-1.1685	4.1746
25	3	0.44415	0.47419	2.266	1.926	-0.77239	7.8746
25	4	0.44744	0.48284	1.618	1.883	-3.2692	5.2119
25	5	0.48236	0.39729	2.201	2.069	-2.1565	6.1822
25	6	0.47019	0.41981	1.778	2.340	1.6286	1.9427
25	7	0.56448	0.35936	1.041	1.591	-1.908	6.7649
25	8	0.5306	0.36061	2.620	2.041	0.34875	0.69134
25	9	0.27153	0.63197	1.489	2.797	-2.8694	7.2513
25	10	0.30709	0.58795	1.929	3.074	-5.4391	11.3762
26	1	0.56713	0.32647	3.249	2.427	-1.3236	6.974
26	2	0.44545	0.41847	2.148	2.587	-0.65419	5.2006
26	3	0.60302	0.30148	2.725	2.130	0.45624	2.426
26	4	0.34221	0.53506	2.661	2.623	-0.27342	3.51
26	5	0.61909	0.2753	2.994	2.111	3.527	-0.35675
26	6	0.40953	0.46672	2.813	2.036	2.1574	0.75165
26	7	0.55617	0.32232	2.747	1.724	1.8334	0.49651
26	8	0.52138	0.3818	1.695	1.581	2.7959	1.6927

List ID	**Word Order**	**SII(Target)**	**SII(Competing)**	**PESQ(Target)**	**PESQ(Competing)**	**SNR(Target)**	**SNR(Competing)**
26	9	0.30647	0.59564	2.111	2.613	-1.6299	6.1581
26	10	0.52327	0.36701	2.795	1.887	-1.991	4.9487
27	1	0.40098	0.46839	1.409	3.136	-6.2689	12.5698
27	2	0.36118	0.53599	1.792	1.789	-2.1491	3.5166
27	3	0.39443	0.49508	1.812	2.100	-1.031	4.752
27	4	0.60135	0.2825	2.994	2.350	-1.5618	4.9656
27	5	0.28662	0.60799	2.153	3.111	-1.4185	7.0917
27	6	0.4051	0.48279	1.544	2.678	-0.53157	7.9084
27	7	0.31153	0.59522	2.320	3.094	0.60733	5.5278
27	8	0.36822	0.53686	2.158	2.951	-1.3048	5.5916
27	9	0.59882	0.29003	2.382	1.875	-2.77	8.2207
27	10	0.51068	0.3728	2.305	2.292	-0.82287	3.5683
10	1	0.37016	0.46697	2.223	2.451	5.8982	-1.9525
10	2	0.56683	0.28457	2.871	1.290	3.0548	1.6928
10	3	0.65869	0.24858	2.869	1.281	6.5555	-2.487
10	4	0.47989	0.3931	2.798	1.865	8.733	-3.5903
10	5	0.50332	0.38759	2.480	1.604	6.6352	-2.69
10	6	0.46186	0.41463	2.511	1.893	13.4157	-4.4062
10	7	0.382	0.49924	2.466	2.179	8.2252	-2.7036
10	8	0.41724	0.46762	2.908	2.492	6.0195	0.46408
10	9	0.40083	0.49386	2.170	2.151	9.2734	-3.982
10	10	0.41894	0.47259	1.744	2.408	0.79679	1.1706
11	1	0.24307	0.65132	2.255	3.141	5.6459	-1.7069
11	2	0.4493	0.44208	2.958	1.448	9.0972	-3.5837
11	3	0.21768	0.70662	1.777	2.006	4.0087	1.8095
11	4	0.62984	0.25334	3.178	2.369	8.373	-2.2788
11	5	0.39025	0.51853	1.738	2.088	-0.37548	1.3196
11	6	0.34641	0.54955	1.568	1.538	0.39932	2.6599
11	7	0.22038	0.66923	2.578	2.932	-1.5591	6.2049
11	8	0.19085	0.70111	2.211	3.155	2.4725	-1.2468
11	9	0.41443	0.48556	1.945	2.621	5.8053	-2.5091
11	10	0.13399	0.77436	2.062	2.280	0.33891	1.5705
12	1	0.26958	0.61826	2.620	2.359	0.61543	0.93667
12	2	0.40622	0.49482	2.114	2.131	0.19442	3.6937
12	3	0.50482	0.41192	2.465	2.080	7.0048	-0.8558
12	4	0.28275	0.62747	2.200	2.472	5.9272	1.417
12	5	0.10525	0.8245	1.653	2.940	5.895	1.639
12	6	0.29713	0.59695	2.182	2.809	1.26	1.7722
12	7	0.58037	0.33854	3.060	1.823	14.3202	-4.7594
12	8	0.32803	0.53312	2.476	2.454	11.2059	-4.8483
12	9	0.40732	0.48356	2.607	2.554	2.5844	1.1415
12	10	0.38708	0.45997	2.520	1.931	4.6194	4.1737
22	1	0.2498	0.65521	2.006	3.352	4.3347	-2.2415
22	2	0.50259	0.40171	2.980	2.065	8.7115	-0.81916
22	3	0.30595	0.59239	2.115	2.634	1.3035	0.2362
22	4	0.58874	0.3039	3.115	1.996	7.2538	-1.5148
22	5	0.23191	0.70255	1.970	3.152	4.0382	-2.0282

List ID	**Word Order**	**SII(Target)**	**SII(Competing)**	**PESQ(Target)**	**PESQ(Competing)**	**SNR(Target)**	**SNR(Competing)**
22	6	0.37844	0.50215	1.861	3.121	0.74061	4.3592
22	7	0.28266	0.64196	2.017	2.442	1.5677	2.4893
22	8	0.46117	0.44594	2.514	1.959	4.0128	-0.59743
22	9	0.51605	0.38294	2.633	2.038	3.5567	1.9461
22	10	0.44028	0.44459	2.839	2.616	3.8439	-1.9608
23	1	0.63987	0.2643	3.316	1.285	13.1556	-5.9321
23	2	0.42028	0.47052	1.930	2.011	4.2211	-2.7712
23	3	0.37101	0.55144	2.122	2.760	8.0472	-3.9333
23	4	0.27936	0.59909	2.771	3.012	13.8082	-2.8318
23	5	0.45047	0.44178	2.322	2.060	5.0229	0.048595
23	6	0.38572	0.50451	1.899	2.498	-1.4224	5.982
23	7	0.34007	0.57057	2.451	2.757	6.8414	-0.89749
23	8	0.44749	0.43823	2.186	0.915	2.3498	2.0505
23	9	0.20675	0.69995	2.069	2.921	4.5446	-2.6872
23	10	0.53311	0.3916	1.498	1.678	7.6835	-1.1347
24	1	0.42614	0.44994	2.618	2.618	6.6271	-3.4861
24	2	0.44737	0.4442	2.407	1.191	5.7257	-0.75246
24	3	0.36649	0.53229	2.035	2.809	1.1634	-0.84443
24	4	0.54815	0.34503	2.691	2.131	8.0629	1.0249
24	5	0.20399	0.70827	2.093	2.990	8.2443	-4.3508
24	6	0.29393	0.57128	1.883	2.944	0.50566	3.1967
24	7	0.35393	0.51082	2.179	2.194	3.3048	0.19477
24	8	0.65012	0.26329	2.689	1.036	9.5809	-0.13332
24	9	0.36112	0.53957	2.027	2.289	1.3404	3.1314
24	10	0.6868	0.24436	2.975	1.887	5.2233	-2.9795
34	1	0.42145	0.46041	2.088	2.840	2.1501	-1.4536
34	2	0.39347	0.50224	2.411	2.609	4.3756	-1.4249
34	3	0.36472	0.50847	2.212	2.107	5.7989	-2.9461
34	4	0.49116	0.38812	2.638	1.802	5.9118	-0.84317
34	5	0.40823	0.45676	0.912	2.568	4.6443	-0.42857
34	6	0.51295	0.37896	1.182	2.621	0.66542	2.1139
34	7	0.25977	0.6593	2.061	2.526	4.6893	0.9043
34	8	0.48527	0.41034	2.364	2.265	3.6657	-0.48466
34	9	0.42402	0.44936	2.324	2.585	4.9659	-2.1482
34	10	0.51823	0.36928	2.882	2.285	9.3919	-5.3218
35	1	0.38314	0.49701	2.789	0.401	6.8247	-0.83095
35	2	0.47243	0.36551	2.878	2.040	0.23708	0.66939
35	3	0.49511	0.36855	3.169	2.017	11.7441	-5.7185
35	4	0.37187	0.52851	2.153	2.897	2.8705	-1.4713
35	5	0.74621	0.14018	3.115	2.307	8.1268	-3.4538
35	6	0.57765	0.32596	2.799	2.052	7.382	-2.0726
35	7	0.3001	0.58721	2.690	2.723	9.2277	-4.7915
35	8	0.78006	0.15792	3.081	2.220	8.1851	-2.1981
35	9	0.59043	0.32015	3.100	1.957	10.4829	-2.4718
35	10	0.54651	0.34743	3.351	1.721	5.8222	0.61093
36	1	0.64091	0.27863	2.432	1.281	11.5389	-2.9531
36	2	0.27168	0.62082	2.267	2.255	9.4186	-5.4439

List ID	**Word Order**	**SII(Target)**	**SII(Competing)**	**PESQ(Target)**	**PESQ(Competing)**	**SNR(Target)**	**SNR(Competing)**
36	3	0.33906	0.53785	3.057	2.895	3.3487	-0.46191
36	4	0.32097	0.5651	1.551	3.198	12.4611	-1.4452
36	5	0.34123	0.54506	1.809	2.918	2.6516	3.8828
36	6	0.47277	0.45265	2.651	2.425	9.113	-1.1948
36	7	0.44572	0.43141	2.663	2.103	9.2256	-4.3106
36	8	0.49077	0.38656	2.613	1.735	13.9184	-5.3275
36	9	0.3873	0.46244	2.363	2.394	6.2434	-3.1175
36	10	0.64923	0.27167	3.007	1.921	6.5933	-2.3655

Table A.2: Frequency of use of each word

List_ID	Word_Order	Target	Frequency	Competing	Frequency
1	1	tam	675.32	nam	4.48
1	2	sek	2.57	kil	14.88
1	3	ker	0.08	tav	0.99
1	4	mil	8.84	dur	62.77
1	5	gel	114.64	bel	17.05
1	6	yağ	62.77	kel	8.88
1	7	sos	5.74	kek	5.82
1	8	sol	140.28	tay	1.68
1	9	şal	2.33	gol	37.43
1	10	poz	9.98	men	7.93
2	1	dur	62.77	tas	5.62
2	2	gam	4.66	boz	6.91
2	3	kıl	15.23	bek	0.67
2	4	pak	2.66	ben	2023.87
2	5	vay	31.45	bay	29.28
2	6	sen	715.79	hal	50.44
2	7	şak	4.28	kez	616.71
2	8	han	24.55	bar	13.73
2	9	jel	3.75	baz	15.13
2	10	sop	0.37	kem	3.55
3	1	san	15.41	saf	55.74
3	2	nur	30.90	gam	4.66
3	3	bal	28.85	sur	5.49
3	4	sim	1.97	fol	0.67
3	5	toz	40.41	pin	0.41
3	6	kın	0.39	çam	18.09
3	7	pek	582.78	kır	19.06
3	8	bas	11.90	şak	4.28
3	9	saz	12.08	kış	72.22
3	10	sit	2.88	tek	921.09

List_ID	Word_Order	Target	Frequency	Competing	Frequency
4	1	her	2991.27		
4	2	kır	19.06		
4	3	zam	16.81		
4	4	hat	20.23		
4	5	sap	10.68		
4	6	kış	72.22		
4	7	yol	512.51		
4	8	çam	18.09		
4	9	kan	171.06		
4	10	fen	23.01		
5	1	çil	3.33		
5	2	gol	37.43		
5	3	dek	118.10		
5	4	kak	0.36		
5	5	yaz	125.75		
5	6	dil	196.93		
5	7	nam	4.48		
5	8	tak	13.99		
5	9	kam	1.01		
5	10	tim	9.71		
6	1	naz	2.57		
6	2	fan	1.16		
6	3	hap	6.39		
6	4	hin	0.83		
6	5	zil	12.75		
6	6	pek	582.78		
6	7	ten	18.13		
6	8	bit	12.87		
6	9	vur	9.41		
6	10	put	1.95		
7	1			yay	12.10
7	2			taş	102.29
7	3			faz	7.48
7	4			sop	0.37
7	5			koz	4.44
7	6			dam	3.14
7	7			sos	5.74
7	8			kaş	6.26
7	9			kak	0.36
7	10			ver	63.81
8	1			sac	2.15
8	2			tim	9.71

List_ID	Word_Order	Target	Frequency	Competing	Frequency
8	3			doz	9.20
8	4			kur	36.54
8	5			yaş	133.98
8	6			bek	0.67
8	7			çay	123.41
8	8			yat	13.60
8	9			san	15.41
8	10			ker	0.08
9	1			yaz	125.75
9	2			yel	6.85
9	3			naz	2.57
9	4			hak	174.02
9	5			pek	582.78
9	6			lal	1.58
9	7			han	24.55
9	8			kaç	214.33
9	9			bor	8.19
9	10			zam	16.81
10	1	zor	289.61	laz	3.77
10	2	ban	1.58	tak	13.99
10	3	gür	13.69	şan	5.98
10	4	kap	9.23	yem	16.83
10	5	gün	1091.06	kas	22.00
10	6	hol	1.24	yoz	1.68
10	7	hem	1112.87	dil	196.93
10	8	tur	29.22	kut	1.09
10	9	gaz	67.13	jel	3.75
10	10	şah	10.50	güm	4.32
11	1	gem	2.31	kıl	15.23
11	2	lor	1.12	fas	3.99
11	3	caz	11.54	zat	7.56
11	4	dut	62.77	kız	351.59
11	5	dal	21.25	kim	307.41
11	6	sac	2.15	has	29.32
11	7	şer	7.04	tut	15.29
11	8	men	7.93	bin	430.05
11	9	sur	5.49	mor	25.24
11	10	yen	6.04	put	1.95
12	1	men	7.93	gün	1091.06
12	2	doz	9.20	pim	0.65
12	3	çit	1.62	tat	14.33
12	4	çin	58.39	sit	2.88

List_ID	Word_Order	Target	Frequency	Competing	Frequency
12	5	cam	46.23	kep	0.77
12	6	şan	5.98	ham	22.42
12	7	bol	118.67	şer	7.04
12	8	kol	38.93	poz	9.98
12	9	yan	144.24	yağ	62.77
12	10	fal	8.76	yar	6.61
13	1	çil	3.33	gaz	67.13
13	2	gol	37.43	sal	2.27
13	3	dek	118.10	mat	3.77
13	4	kak	0.36	çim	7.44
13	5	yaz	125.75	lor	1.12
13	6	dil	196.93	gül	105.61
13	7	nam	4.48	son	1131.37
13	8	tak	13.99	var	1830.37
13	9	kam	1.01	kaz	6.08
13	10	tim	9.71	fen	23.01
14	1	naz	2.57	laz	3.77
14	2	fan	1.16	tak	13.99
14	3	hap	6.39	şan	5.98
14	4	hin	0.83	yem	16.83
14	5	zil	12.75	kas	22.00
14	6	pek	582.78	yoz	1.68
14	7	ten	18.13	dil	196.93
14	8	bit	12.87	kut	1.09
14	9	vur	9.41	jel	3.75
14	10	put	1.95	güm	4.32
15	1	her	2991.27	yay	12.10
15	2	kır	19.06	taş	102.29
15	3	zam	16.81	faz	7.48
15	4	hat	20.23	sop	0.37
15	5	sap	10.68	koz	4.44
15	6	kış	72.22	dam	3.14
15	7	yol	512.51	sos	5.74
15	8	çam	18.09	kaş	6.26
15	9	kan	171.06	kak	0.36
15	10	fen	23.01	ver	63.81
16	1	san	15.41		
16	2	nur	30.90		
16	3	bal	28.85		
16	4	sim	1.97		
16	5	toz	40.41		
16	6	kın	0.39		

List_ID	Word_Order	Target	Frequency	Competing	Frequency
16	7	pek	582.78		
16	8	bas	11.90		
16	9	saz	12.08		
16	10	sit	2.88		
17	1	men	7.93		
17	2	doz	9.20		
17	3	çit	1.62		
17	4	çin	58.39		
17	5	cam	46.23		
17	6	şan	5.98		
17	7	bol	118.67		
17	8	kol	38.93		
17	9	yan	144.24		
17	10	fal	8.76		
18	1	hit	1.95		
18	2	keş	1.11		
18	3	cin	13.44		
18	4	hay	19.91		
18	5	zar	16.87		
18	6	baş	149.26		
18	7	haz	20.46		
18	8	kaş	6.26		
18	9	güz	9.89		
18	10	bul	9.67		
19	1			nam	4.48
19	2			kil	14.88
19	3			tav	0.99
19	4			dur	62.77
19	5			bel	17.05
19	6			kel	8.88
19	7			kek	5.82
19	8			tay	1.68
19	9			gol	37.43
19	10			men	7.93
20	1			saf	55.74
20	2			gam	4.66
20	3			sur	5.49
20	4			fol	0.67
20	5			pin	0.41
20	6			çam	18.09
20	7			kır	19.06
20	8			şak	4.28

List_ID	Word_Order	Target	Frequency	Competing	Frequency
20	9			kış	72.22
20	10			tek	921.09
21	1			tas	5.62
21	2			boz	6.91
21	3			bek	0.67
21	4			ben	2023.87
21	5			bay	29.28
21	6			hal	50.44
21	7			kez	616.71
21	8			bar	13.73
21	9			baz	15.13
21	10			kem	3.55
22	1	gem	2.31	yay	12.10
22	2	lor	1.12	taş	102.29
22	3	caz	11.54	faz	7.48
22	4	dut	62.77	sop	0.37
22	5	dal	21.25	koz	4.44
22	6	sac	2.15	dam	3.14
22	7	şer	7.04	sos	5.74
22	8	men	7.93	kaş	6.26
22	9	sur	5.49	kak	0.36
22	10	yen	6.04	ver	63.81
23	1	dur	62.77	sac	2.15
23	2	gam	4.66	tim	9.71
23	3	kıl	15.23	doz	9.20
23	4	pak	2.66	kur	36.54
23	5	vay	31.45	yaş	133.98
23	6	sen	715.79	bek	0.67
23	7	şak	4.28	çay	123.41
23	8	han	24.55	yat	13.60
23	9	jel	3.75	san	15.41
23	10	sop	0.37	ker	0.08
24	1	tam	675.32	gaz	67.13
24	2	sek	2.57	sal	2.27
24	3	ker	0.08	mat	3.77
24	4	mil	8.84	çim	7.44
24	5	gel	114.64	lor	1.12
24	6	yağ	62.77	gül	105.61
24	7	sos	5.74	son	1131.37
24	8	sol	140.28	var	1830.37
24	9	şal	2.33	kaz	6.08
24	10	poz	9.98	fen	23.01

List_ID	Word_Order	Target	Frequency	Competing	Frequency
25	1	hit	1.95	kıl	15.23
25	2	keş	1.11	fas	3.99
25	3	cin	13.44	mat	3.77
25	4	hay	19.91	çim	7.44
25	5	zar	16.87	lor	1.12
25	6	baş	149.26	gül	105.61
25	7	haz	20.46	son	1131.37
25	8	kaş	6.26	var	1830.37
25	9	güz	9.89	kaz	6.08
25	10	bul	9.67	fen	23.01
26	1	gem	2.31	yaz	125.75
26	2	lor	1.12	yel	6.85
26	3	caz	11.54	naz	2.57
26	4	dut	62.77	hak	174.02
26	5	dal	21.25	pek	582.78
26	6	sac	2.15	lal	1.58
26	7	şer	7.04	han	24.55
26	8	men	7.93	kaç	214.33
26	9	sur	5.49	bor	8.19
26	10	yen	6.04	zam	16.81
27	1	zor	289.61	sac	2.15
27	2	ban	1.58	tim	9.71
27	3	gür	13.69	doz	9.20
27	4	kap	9.23	kur	36.54
27	5	gün	1091.06	yaş	133.98
27	6	hol	1.24	bek	0.67
27	7	hem	1112.87	çay	123.41
27	8	tur	29.22	yat	13.60
27	9	gaz	67.13	san	15.41
27	10	şah	10.50	ker	0.08
28	1	dur	62.77		
28	2	gam	4.66		
28	3	kıl	15.23		
28	4	pak	2.66		
28	5	vay	31.45		
28	6	sen	715.79		
28	7	şak	4.28		
28	8	han	24.55		
28	9	jel	3.75		
28	10	sop	0.37		
29	1	tam	675.32		
29	2	sek	2.57		

List_ID	Word_Order	Target	Frequency	Competing	Frequency
29	3	ker	0.08		
29	4	mil	8.84		
29	5	gel	114.64		
29	6	yağ	62.77		
29	7	sos	5.74		
29	8	sol	140.28		
29	9	şal	2.33		
29	10	poz	9.98		
30	1	gem	2.31		
30	2	lor	1.12		
30	3	caz	11.54		
30	4	dut	62.77		
30	5	dal	21.25		
30	6	sac	2.15		
30	7	şer	7.04		
30	8	men	7.93		
30	9	sur	5.49		
30	10	yen	6.04		
31	1			gün	1091.06
31	2			pim	0.65
31	3			tat	14.33
31	4			sit	2.88
31	5			kep	0.77
31	6			ham	22.42
31	7			şer	7.04
31	8			poz	9.98
31	9			yağ	62.77
31	10			yar	6.61
32	1			gaz	67.13
32	2			sal	2.27
32	3			mat	3.77
32	4			çim	7.44
32	5			lor	1.12
32	6			gül	105.61
32	7			son	1131.37
32	8			var	1830.37
32	9			kaz	6.08
32	10			fen	23.01
33	1			laz	3.77
33	2			tak	13.99
33	3			şan	5.98
33	4			yem	16.83

List_ID	Word_Order	Target	Frequency	Competing	Frequency
33	5			kas	22.00
33	6			yoz	1.68
33	7			dil	196.93
33	8			kut	1.09
33	9			jel	3.75
33	10			güm	4.32
34	1	her	2991.27	nam	4.48
34	2	kır	19.06	kil	14.88
34	3	zam	16.81	tav	0.99
34	4	hat	20.23	dur	62.77
34	5	sap	10.68	bel	17.05
34	6	kış	72.22	kel	8.88
34	7	yol	512.51	kek	5.82
34	8	çam	18.09	tay	1.68
34	9	kan	171.06	gol	37.43
34	10	fen	23.01	men	7.93
35	1	naz	2.57	saf	55.74
35	2	fan	1.16	gam	4.66
35	3	hap	6.39	sur	5.49
35	4	hin	0.83	fol	0.67
35	5	zil	12.75	pin	0.41
35	6	pek	582.78	çam	18.09
35	7	ten	18.13	kır	19.06
35	8	bit	12.87	şak	4.28
35	9	vur	9.41	kış	72.22
35	10	put	1.95	tek	921.09
36	1	çil	3.33	tas	5.62
36	2	gol	37.43	boz	6.91
36	3	dek	118.10	bek	0.67
36	4	kak	0.36	ben	2023.87
36	5	yaz	125.75	bay	29.28
36	6	dil	196.93	hal	50.44
36	7	nam	4.48	kez	616.71
36	8	tak	13.99	bar	13.73
36	9	kam	1.01	baz	15.13
36	10	tim	9.71	kem	3.55

Table A.3: Sentiment analysis of stimuli

List_ID	Word	Tone	Polarity
1	tam	0,148966115	1
1	sek	-0,24318	-1
1	ker	NA	NA

List_ID	Word	Tone	Polarity
1	mil	-0,066297456	-1
1	gel	0,11589834	1
1	yak	-0,29213	-1
1	sos	0,183785303	1
1	sol	-0,183337576	-1
1	şad	0,189079786	1
1	poz	-0,027873912	-1
1	nam	0,071370339	1
1	kil	-0,258555639	-1
1	tav	-0,147548909	-1
1	dur	-0,044949748	-1
1	bel	-0,08742118	-1
1	kel	-0,590968344	-1
1	kek	-0,043424825	-1
1	tay	0,103438748	1
1	gol	0,085687439	1
1	men	-0,169282134	-1
2	dur	-0,044949748	-1
2	gam	-0,071825297	-1
2	kıl	0,097675054	1
2	pak	0,125258028	1
2	vay	0,056867752	1
2	sen	-0,307077215	-1
2	şak	-0,43104	-1
2	han	-0,031584543	-1
2	jel	-0,4899	-1
2	sop	0,183907381	1
2	tas	0,042581184	1
2	boz	0,296751685	1
2	bek	0,059547439	1
2	ben	0,083792955	1
2	bay	0,083206913	1
2	hal	0,078325365	1
2	kez	0,039976547	1
2	bar	-0,050879814	-1
2	baz	-0,14492	-1
2	kem	-0,140702155	-1
3	san	0,066213254	1
3	nur	0,055558566	1
3	bal	0,078187439	1
3	sim	-0,042778688	-1
3	toz	-0,051572279	-1
3	kın	-0,050198735	-1
3	pek	0,080607208	1
3	bas	-0,158278609	-1
3	saz	0,134870186	1
3	sit	NA	NA
3	saf	-0,161642144	-1
3	gam	-0,071825297	-1
3	sur	-0,126358719	-1
3	fol	0,191477439	1
3	pin	NA	NA
3	çam	0,074007439	1
3	kır	-0,157658747	-1

List_ID	Word	Tone	Polarity
3	şak	-0,43104	-1
3	kış	-0,029927846	-1
3	tek	0,083084469	1
4	her	0,120672619	1
4	kır	-0,157658747	-1
4	zam	-0,047365176	-1
4	hat	0,043237359	1
4	sap	-0,063190337	-1
4	kış	-0,029927846	-1
4	yol	-0,057189	-1
4	çan	0,140117341	1
4	kan	-0,086161002	-1
4	fen	0,073748735	1
5	çil	0,042037874	1
5	gol	0,085687439	1
5	dek	-0,074059989	-1
5	kak	-0,330989541	-1
5	yas	-0,236873856	-1
5	dil	-0,011273306	-1
5	nam	0,071370339	1
5	tak	0,091117439	1
5	kam	-0,036318639	-1
5	tim	-0,085394809	-1
6	naz	0,286694973	1
6	fan	0,098498652	1
6	hap	-0,139559462	-1
6	hin	-0,028435015	-1
6	zil	0,183878553	1
6	tek	0,083084469	1
6	ten	0,078488796	1
6	bit	-0,477937319	-1
6	vur	-0,24067	-1
6	put	0,146535534	1
7	yay	-0,062323238	-1
7	taş	0,077185313	1
7	faz	0,048457914	1
7	sop	0,183907381	1
7	koz	-0,0597085	-1
7	dam	-0,14672353	-1
7	sos	0,183785303	1
7	kaş	0,118973983	1
7	kak	-0,330989541	-1
7	ver	0,047196629	1
8	sac	0,044357439	1
8	tim	-0,085394809	-1
8	doz	0,059414211	1
8	kur	0,04026721	1
8	yaş	0,100765048	1
8	bek	0,059547439	1
8	çay	0,037999447	1
8	yat	-0,017146635	-1
8	san	0,066213254	1
8	ker	NA	NA
9	yaz	0,123932313	1

List_ID	Word	Tone	Polarity
9	yel	0,071679329	1
9	naz	0,286694973	1
9	hak	0,235589339	1
9	pek	0,080607208	1
9	lal	-0,220320623	-1
9	han	-0,031584543	-1
9	kaç	-0,29641	-1
9	bor	-0,167391475	-1
9	zam	-0,047365176	-1
10	zor	-0,189052131	-1
10	ban	-0,035117682	-1
10	gür	0,070920025	1
10	kap	-0,08042855	-1
10	gün	-0,002912194	-1
10	hol	-0,100955902	-1
10	hem	0,210301714	1
10	tur	0,088820684	1
10	gaz	-0,038324601	-1
10	şap	0,118187439	1
10	laz	NA	NA
10	tak	0,091117439	1
10	şan	0,145344696	1
10	yem	0,055143151	1
10	kas	-0,092588234	-1
10	yoz	-0,202761737	-1
10	dil	-0,011273306	-1
10	kut	0,100706834	1
10	jel	-0,4899	-1
10	güm	-0,03409	-1
11	gem	-0,23851	-1
11	lor	0,283947439	1
11	caz	0,135175368	1
11	dut	0,152612439	1
11	dal	0,094159182	1
11	sac	0,044357439	1
11	şer	-0,008592358	-1
11	men	-0,169282134	-1
11	sur	-0,126358719	-1
11	yen	0,118496352	1
11	kıl	0,097675054	1
11	fas	NA	NA
11	zat	-0,036549428	-1
11	kız	-0,048211899	-1
11	kim	-0,218678	-1
11	has	-0,038224636	-1
11	tut	0,054767439	1
11	bin	0,148619891	1
11	mor	0,094347689	1
11	put	0,146535534	1
12	nem	0,050040866	1
12	doz	0,059414211	1
12	çit	0,101379449	1
12	çin	-0,04417	-1
12	cam	-0,007547864	-1

List_ID	Word	Tone	Polarity
12	şan	0,145344696	1
12	bol	0,079168918	1
12	kol	-0,04430723	-1
12	yan	-0,102625909	-1
12	fal	0,099395149	1
12	gün	-0,002912194	-1
12	pim	0,120887439	1
12	tat	0,140723134	1
12	sit	NA	NA
12	kep	-0,108842477	-1
12	ham	-0,205823412	-1
12	şer	-0,008592358	-1
12	poz	-0,027873912	-1
12	yağ	-0,08682	-1
12	yar	0,05247246	1
13	çil	0,042037874	1
13	gol	0,085687439	1
13	dek	-0,074059989	-1
13	kak	-0,330989541	-1
13	yas	-0,236873856	-1
13	dil	-0,011273306	-1
13	nam	0,071370339	1
13	tak	0,091117439	1
13	kam	-0,036318639	-1
13	tim	-0,085394809	-1
13	gaz	-0,038324601	-1
13	sal	0,123182629	1
13	mat	-0,099227153	-1
13	çim	-0,099983209	-1
13	lor	0,283947439	1
13	gül	0,061939939	1
13	son	-0,07265704	-1
13	var	0,048372451	1
13	kaz	-0,121679957	-1
13	fen	0,073748735	1
14	naz	0,286694973	1
14	fan	0,098498652	1
14	hap	-0,139559462	-1
14	hin	-0,028435015	-1
14	zil	0,183878553	1
14	tek	0,083084469	1
14	ten	0,078488796	1
14	bit	-0,477937319	-1
14	vur	-0,24067	-1
14	put	0,146535534	1
14	laz	NA	NA
14	tak	0,091117439	1
14	şan	0,145344696	1
14	yem	0,055143151	1
14	kas	-0,092588234	-1
14	yoz	-0,202761737	-1
14	dil	-0,011273306	-1
14	kut	0,100706834	1
14	jel	-0,4899	-1

List_ID	Word	Tone	Polarity
14	güm	-0,03409	-1
15	her	0,120672619	1
15	kır	-0,157658747	-1
15	zam	-0,047365176	-1
15	hat	0,043237359	1
15	sap	-0,063190337	-1
15	kış	-0,029927846	-1
15	yol	-0,057189	-1
15	çan	0,140117341	1
15	kan	-0,086161002	-1
15	fen	0,073748735	1
15	yay	-0,062323238	-1
15	taş	0,077185313	1
15	faz	0,048457914	1
15	sop	0,183907381	1
15	koz	-0,0597085	-1
15	dam	-0,14672353	-1
15	sos	0,183785303	1
15	kaş	0,118973983	1
15	kak	-0,330989541	-1
15	ver	0,047196629	1
16	san	0,066213254	1
16	nur	0,055558566	1
16	bal	0,078187439	1
16	sim	-0,042778688	-1
16	toz	-0,051572279	-1
16	kın	-0,050198735	-1
16	pek	0,080607208	1
16	bas	-0,158278609	-1
16	saz	0,134870186	1
16	sit	NA	NA
17	nem	0,050040866	1
17	doz	0,059414211	1
17	çit	0,101379449	1
17	çin	-0,04417	-1
17	cam	-0,007547864	-1
17	şan	0,145344696	1
17	bol	0,079168918	1
17	kol	-0,04430723	-1
17	yan	-0,102625909	-1
17	fal	0,099395149	1
18	hit	NA	NA
18	keş	0,276057439	1
18	cin	-0,129106767	-1
18	hay	0,087697439	1
18	zar	-0,000547085	-1
18	baş	0,094220131	1
18	haz	0,251695494	1
18	kaş	0,118973983	1
18	güz	0,106287105	1
18	bul	0,109192827	1
19	nam	0,071370339	1
19	kil	-0,258555639	-1
19	tav	-0,147548909	-1

List_ID	Word	Tone	Polarity
19	dur	-0,044949748	-1
19	bel	-0,08742118	-1
19	kel	-0,590968344	-1
19	kek	-0,043424825	-1
19	tay	0,103438748	1
19	gol	0,085687439	1
19	men	-0,169282134	-1
20	saf	-0,161642144	-1
20	gam	-0,071825297	-1
20	sur	-0,126358719	-1
20	fol	0,191477439	1
20	pin	NA	NA
20	çam	0,074007439	1
20	kır	-0,157658747	-1
20	şak	-0,43104	-1
20	kış	-0,029927846	-1
20	tek	0,083084469	1
21	tas	0,042581184	1
21	boz	0,296751685	1
21	bek	0,059547439	1
21	ben	0,083792955	1
21	bay	0,083206913	1
21	hal	0,078325365	1
21	kez	0,039976547	1
21	bar	-0,050879814	-1
21	baz	-0,14492	-1
21	kem	-0,140702155	-1
22	gem	-0,23851	-1
22	lor	0,283947439	1
22	caz	0,135175368	1
22	dut	0,152612439	1
22	dal	0,094159182	1
22	sac	0,044357439	1
22	şer	-0,008592358	-1
22	men	-0,169282134	-1
22	sur	-0,126358719	-1
22	yen	0,118496352	1
22	yay	-0,062323238	-1
22	taş	0,077185313	1
22	faz	0,048457914	1
22	sop	0,183907381	1
22	koz	-0,0597085	-1
22	dam	-0,14672353	-1
22	sos	0,183785303	1
22	kaş	0,118973983	1
22	kak	-0,330989541	-1
22	ver	0,047196629	1
23	dur	-0,044949748	-1
23	gam	-0,071825297	-1
23	kıl	0,097675054	1
23	pak	0,125258028	1
23	vay	0,056867752	1
23	sen	-0,307077215	-1
23	şak	-0,43104	-1

List_ID	Word	Tone	Polarity
23	han	-0,031584543	-1
23	jel	-0,4899	-1
23	sop	0,183907381	1
23	sac	0,044357439	1
23	tim	-0,085394809	-1
23	doz	0,059414211	1
23	kur	0,04026721	1
23	yaş	0,100765048	1
23	bek	0,059547439	1
23	çay	0,037999447	1
23	yat	-0,017146635	-1
23	san	0,066213254	1
23	ker	NA	NA
24	tam	0,148966115	1
24	sek	-0,24318	-1
24	ker	NA	NA
24	mil	-0,066297456	-1
24	gel	0,11589834	1
24	yak	-0,29213	-1
24	sos	0,183785303	1
24	sol	-0,183337576	-1
24	şad	0,189079786	1
24	poz	-0,027873912	-1
24	gaz	-0,038324601	-1
24	sal	0,123182629	1
24	mat	-0,099227153	-1
24	çim	-0,099983209	-1
24	lor	0,283947439	1
24	gül	0,061939939	1
24	son	-0,07265704	-1
24	var	0,048372451	1
24	kaz	-0,121679957	-1
24	fen	0,073748735	1
25	hit	NA	NA
25	keş	0,276057439	1
25	cin	-0,129106767	-1
25	hay	0,087697439	1
25	zar	-0,000547085	-1
25	baş	0,094220131	1
25	haz	0,251695494	1
25	kaş	0,118973983	1
25	güz	0,106287105	1
25	bul	0,109192827	1
25	kıl	0,097675054	1
25	fas	NA	NA
25	zat	-0,036549428	-1
25	kız	-0,048211899	-1
25	kim	-0,218678	-1
25	has	-0,038224636	-1
25	tut	0,054767439	1
25	bin	0,148619891	1
25	mor	0,094347689	1
25	put	0,146535534	1
26	gem	-0,23851	-1

List_ID	Word	Tone	Polarity
26	lor	0,283947439	1
26	caz	0,135175368	1
26	dut	0,152612439	1
26	dal	0,094159182	1
26	sac	0,044357439	1
26	şer	-0,008592358	-1
26	men	-0,169282134	-1
26	sur	-0,126358719	-1
26	yen	0,118496352	1
26	yaz	0,123932313	1
26	yel	0,071679329	1
26	naz	0,286694973	1
26	hak	0,235589339	1
26	pek	0,080607208	1
26	lal	-0,220320623	-1
26	han	-0,031584543	-1
26	kaç	-0,29641	-1
26	bor	-0,167391475	-1
26	zam	-0,047365176	-1
27	zor	-0,189052131	-1
27	ban	-0,035117682	-1
27	gür	0,070920025	1
27	kap	-0,08042855	-1
27	gün	-0,002912194	-1
27	hol	-0,100955902	-1
27	hem	0,210301714	1
27	tur	0,088820684	1
27	gaz	-0,038324601	-1
27	şap	0,118187439	1
27	sac	0,044357439	1
27	tim	-0,085394809	-1
27	doz	0,059414211	1
27	kur	0,04026721	1
27	yaş	0,100765048	1
27	bek	0,059547439	1
27	çay	0,037999447	1
27	yat	-0,017146635	-1
27	san	0,066213254	1
27	ker	NA	NA
28	dur	-0,044949748	-1
28	gam	-0,071825297	-1
28	kıl	0,097675054	1
28	pak	0,125258028	1
28	vay	0,056867752	1
28	sen	-0,307077215	-1
28	şak	-0,43104	-1
28	han	-0,031584543	-1
28	jel	-0,4899	-1
28	sop	0,183907381	1
29	tam	0,148966115	1
29	sek	-0,24318	-1
29	ker	NA	NA
29	mil	-0,066297456	-1
29	gel	0,11589834	1

List_ID	Word	Tone	Polarity
29	yak	-0,29213	-1
29	sos	0,183785303	1
29	sol	-0,183337576	-1
29	şad	0,189079786	1
29	poz	-0,027873912	-1
30	gem	-0,23851	-1
30	lor	0,283947439	1
30	caz	0,135175368	1
30	dut	0,152612439	1
30	dal	0,094159182	1
30	sac	0,044357439	1
30	şer	-0,008592358	-1
30	men	-0,169282134	-1
30	sur	-0,126358719	-1
30	yen	0,118496352	1
31	gün	-0,002912194	-1
31	pim	0,120887439	1
31	tat	0,140723134	1
31	sit	NA	NA
31	kep	-0,108842477	-1
31	ham	-0,205823412	-1
31	şer	-0,008592358	-1
31	poz	-0,027873912	-1
31	yağ	-0,08682	-1
31	yar	0,05247246	1
32	gaz	-0,038324601	-1
32	sal	0,123182629	1
32	mat	-0,099227153	-1
32	çim	-0,099983209	-1
32	lor	0,283947439	1
32	gül	0,061939939	1
32	son	-0,07265704	-1
32	var	0,048372451	1
32	kaz	-0,121679957	-1
32	fen	0,073748735	1
33	laz	NA	NA
33	tak	0,091117439	1
33	şan	0,145344696	1
33	yem	0,055143151	1
33	kas	-0,092588234	-1
33	yoz	-0,202761737	-1
33	dil	-0,011273306	-1
33	kut	0,100706834	1
33	jel	-0,4899	-1
33	güm	-0,03409	-1
34	her	0,120672619	1
34	kır	-0,157658747	-1
34	zam	-0,047365176	-1
34	hat	0,043237359	1
34	sap	-0,063190337	-1
34	kış	-0,029927846	-1
34	yol	-0,057189	-1
34	çan	0,140117341	1
34	kan	-0,086161002	-1

List_ID	Word	Tone	Polarity
34	fen	0,073748735	1
34	nam	0,071370339	1
34	kil	-0,258555639	-1
34	tav	-0,147548909	-1
34	dur	-0,044949748	-1
34	bel	-0,08742118	-1
34	kel	-0,590968344	-1
34	kek	-0,043424825	-1
34	tay	0,103438748	1
34	gol	0,085687439	1
34	men	-0,169282134	-1
35	naz	0,286694973	1
35	fan	0,098498652	1
35	hap	-0,139559462	-1
35	hin	-0,028435015	-1
35	zil	0,183878553	1
35	tek	0,083084469	1
35	ten	0,078488796	1
35	bit	-0,477937319	-1
35	vur	-0,24067	-1
35	put	0,146535534	1
35	saf	-0,161642144	-1
35	gam	-0,071825297	-1
35	sur	-0,126358719	-1
35	fol	0,191477439	1
35	pin	NA	NA
35	çam	0,074007439	1
35	kır	-0,157658747	-1
35	şak	-0,43104	-1
35	kış	-0,029927846	-1
35	tek	0,083084469	1
36	çil	0,042037874	1
36	gol	0,085687439	1
36	dek	-0,074059989	-1
36	kak	-0,330989541	-1
36	yas	-0,236873856	-1
36	dil	-0,011273306	-1
36	nam	0,071370339	1
36	tak	0,091117439	1
36	kam	-0,036318639	-1
36	tim	-0,085394809	-1
36	tas	0,042581184	1
36	boz	0,296751685	1
36	bek	0,059547439	1
36	ben	0,083792955	1
36	bay	0,083206913	1
36	hal	0,078325365	1
36	kez	0,039976547	1
36	bar	-0,050879814	-1
36	baz	-0,14492	-1
36	kem	-0,140702155	-1

7.2. APPENDIX B

UYGULAMALI ETİK ARAŞTIRMA MERKEZİ
APPLIED ETHICS RESEARCH CENTER

ORTA DOĞU TEKNİK ÜNİVERSİTESİ
MIDDLE EAST TECHNICAL UNIVERSITY

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

Sayı: 28620816 / 60

20 Şubat 2019

Konu: Değerlendirme Sonucu

Gönderen: ODTÜ İnsan Araştırmaları Etik Kurulu (İAEK)

İlgi: İnsan Araştırmaları Etik Kurulu Başvurusu

Sayın Doç.Dr. Banu Günel KILIÇ

Danışmanlığını yaptığınız ***Serap YAĞMUR' un*** **"Değişik Çevre Ortamlarında Dinleme Çabasının Ölçülmesi ve Modellenmesi"** başlıklı araştırması İnsan Araştırmaları Etik Kurulu tarafından uygun görülmüş ve **051-ODTÜ-2019** protokol numarası ile onaylanmıştır.

Saygılarımızla bilgilerinize sunarız.

CIRRICULUM VITAE

SERAP YAĞMUR

Information System Department

Middle East Technical University

Ankara, Turkey

AREAS OF INTEREST

Eye Tracker Methodology, Listening Effort Measurement and Modeling, Auditory Attention, Speech Intelligibility, Human Computer Interaction, Usability, Mobile Learning, E-Learning, Distance Education, Computer Supported Collaborative Learning, Robotics Programming.

EDUCATION

Middle East Technical University, Ankara, Turkey
2014-2023
Doctor Philosophy in Information Systems
Thesis Title: Listening Effort Measurement and Modeling for Different Environmental Conditions
Supervisor: Prof. Dr. Banu Günel Kılıç
Co-Advisor: Prof. Dr. Kürşat Çağıltay

Exchange Student, Non-Europe Erasmus, Software College, Northeastern University, Shenyang/China
March-June 2018
Advisor: Prof. Dr. Yao Yudong
Research Subject: Deep Learning Applications and Practices

Visiting Student, Department of Informatics, University of Hamburg, Germany
June-October 2017
Project Title: “Civitas Digitalis” Smart Cities
Advisor: Prof. Dr. Eva Bittner
Research Subject: Literature Review on Chatbots

Anadolu University, Ankara, Turkey
2018-2020
Open Education Faculty
Associate degree in Department of Justice

Hacettepe University, Ankara, Turkey
2014-2018
Institute of Social Sciences
Master on Law in Information Law (Non-Thesis)

Middle East Technical University, Ankara, Turkey
2009-2014
Master of Science in Information Systems
Thesis: Usability Evaluation of Dynamic Geometry Software Through Eye Tracking and Communication Breakdown Analysis
Supervisor: Assoc. Prof. Dr. Murat Perit Çakır

Atatürk University, Erzurum, Turkey
2005-2009
Bachelor of Computer Education and Instructional Technology

TEACHING ASSISTANTSHIP

- IS 535 Regulatory and Legal Aspects of Information Systems, Assistant
- IS 573 Human Computer Interaction, Assistant
- IS 585 Social Network Analysis, Assistant
- IS 737 Contemporary Socio-technical Perspectives in Information Systems, Assistant
- IS 739 Information Systems in Organizational Design and Applied Systems Thinking, Assistant
- IS 740 Information Technology Acceptance in Organizations, Assistant
- IS 771 Security Engineering, Assistant
- SM 514 Introduction to Software Testing, Assistant
- SM 547 Model Based Systems Engineering, Assistant
- IS 100 Introduction to Information Technologies and Applications, Instructor

LIST OF PUBLICATIONS

Yağmur, S. Kılıç, B.G., Çağıltay, K. (Under Reviewer, 2023). Effects of Selective Directional Auditory Attention on Eye Movements and Pupil Responses.

Yağmur, S., & Çakır, M. P. (2016, July). Usability Evaluation of a Dynamic Geometry Software Mobile Interface Through Eye Tracking. In *International Conference on Learning and Collaboration Technologies* (pp. 391-402). Springer International Publishing.

Yağmur. S. (2016). A Literature Review: Usability Aspects of Ubiquitous Computing, *Platcon16, 2016 International Conference on Platform Technology and Service.*

Yağmur, S. (2015). A Survey on Usability Aspects of Ubiquitous Computing. *Journal of Platform Technology, 3(3), 22-34.*

Yagmur, S. Cagiltay K. (2013). A Usability Study of Dynamic Geometry Software's Interfaces, *HCI International 2013 - Posters' Extended Abstracts*

Yagmur, S. (2012). Podcast Literature Review, *CSEIT 2012: 3rd Annual International Conference*