EFFECTS OF VOICE FAMILIARITY ON AUDITORY DISTANCE PERCEPTION

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

ÖZGEN DEMİRKAPLAN

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

JUNE 2017

Approval of the thesis:

EFFECTS OF VOICE FAMILIARITY ON AUDITORY DISTANCE PERCEPTION

submitted by **ÖZGEN DEMİRKAPLAN** in partial fulfillment of the requirements for the degree of **Master of Science in Cognitive Science Department, Middle East Technical University** by,

Prof. Dr. Deniz Zeyrek Bozşahin
Director, Graduate School of **Institute of Informatics** ________________

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

Assoc. Prof. Dr. Hüseyin Hacıhabiboğlu
Supervisor, **ModSim, METU** ________________

Examining Committee Members:

Prof. Cem Bozşahin
Cognitive Science, METU ________________

Assoc. Prof. Hüseyin Hacıhabiboğlu
Multimedia Informatics, METU ________________

Prof. Deniz Zeyrek Bozşahin
Cognitive Science, METU ________________

Asst. Prof. Murat Perit Çakır
Cognitive Science, METU ________________

Asst. Prof. Hacı Hulusi Kafalıgönül
Neuroscience, Bilkent University ________________

Date: ________________

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

Name, Last Name: ÖZGEN DEMİRKAPLAN

Signature :

ABSTRACT

EFFECTS OF VOICE FAMILIARITY ON AUDITORY DISTANCE PERCEPTION

DEMİRKAPLAN, Özgen

M.S., Department of Cognitive Science

Supervisor : Assoc. Prof. Dr. Hüseyin Hacıhabiboğlu

JUNE 2017, 57 pages

Auditory distance perception is a multidimensional phenomenon. Familiarity with the sound source is known to have an important effect on the distance perception as one of the cognitive cues. An auditory distance perception experiment to assess the effects of interpersonal familiarity on auditory distance perception is reported in this article. The subjective experiment involves a binaural listening experiment where different source distances between 0.5 and 16 meters were simulated. The participants are 12 heterosexual couples who have known each other for at least two years with daily interaction. All couples were strangers to other participated couples. Each subject listened and judged the distance of five different speech utterances from their partner (F0), the spectrally most similar stranger (F1) and the spectrally most dissimilar stranger (F2). The results indicate that there is no significant effect of interpersonal familiarity on auditory distance perception for the three conditions of familiarity. For further investigation, results revealed another interesting point that other cognitive factors besides the interpersonal familiarity (e.g. semantic aspects of the utterances listened) could be as useful as acoustic cues on the enhancement of auditory distance perception.

Keywords: auditory familiarity, auditory distance perception

ÖZ

KONUŞMA SESİNDE AŞİNALIĞININ İŞİTSEL UZAKLIK ALGISINA ETKİLERİ

DEMİRKAPLAN, Özgen
Yüksek Lisans, Bilişsel Bilimler Bölümü
Tez Yöneticisi : Doç. Dr. Hüseyin Hacıhabiboğlu

Haziran 2017 , 57 sayfa

İşitsel uzaklık algısı çok boyutlu bir olgudur. Ses kaynağına olan aşinalığın, bilişsel ipuçlarından biri olarak mesafe algısında önemli bir etkiye sahip olduğu bilinmektedir. Bu çalışmada, kişiler arası yakınlığın işitsel uzaklık algısı üzerindeki etkilerini değerlendirmek için öznel bir işitsel uzaklık algılama deneyi sunulmaktadır. Öznel deney, 0.5 ve 16 metre arasındaki farklı ses kaynaklarının mesafelerinin simüle edildiği binaural bir dinleme içermekteydi. Katılımcılar 12, farklı cinsiyetten çiftlerdi. Birbirlerini en az bir yıldır tanıyan bu çiftler, kendileri dışındaki diğer çiftlere yabancıydılar. Her bir kişi, eşinden (F0), eşinin sesine en benzeyen yabancıdan (F1) ve eşinin sesine en farklı olan yabancıdan (F2) gelen 5 farklı konuşma kaydını dinleyerek uzaklıklarını değerlendirmiştir. Elde edilen sonuçlara göre, aşinalığın bu üç seviyesi için aşinalığın işitsel uzaklık algısında belirgin bir etkisinin olmadığını gözlenmiştir. Daha detaylı bir çalışmayı gerektiren bir sonuç olarak, kişiler arası ses aşinalığının yanı sıra diğer bilişsel faktörlerin (örneğin, cümlelerin semantik veya fonetik özelliği), işitsel uzaklık algısının geliştirilmesinde en az akustik ipuçları kadar etkili olabileceğini işaret etmiştir.

Anahtar Kelimeler: işitsel aşinalık, işitsel uzaklık algısı

I dedicate my dissertation to my family and to all of my friends. A special feeling of gratitude to my loving parents, Cevdet and Selma Demirkaplan, whose support was endless and keep me going forward all the time.

My beloved friends; Ayşenur Hülagü, Sinan Onur Altınuç and Umutcan İpekoğlu whose contributions to my study both emotionally and physically was priceless.

ACKNOWLEDGMENTS

I would like to express the deepest appreciation to my committee member and my thesis advisor Associate Professor Hüseyin Hacıhabiboğlu, who has the attitude and the substance of a genius: he continually and convincingly conveyed a spirit of adventure in regards to research and academic writing and an excitement in regard to teaching. Without his guidance and persistent help, this dissertation would not have been possible. Also, his notions and helps on my study demonstrated to me that the importance of academic ethics, relevant criticism and reasoning, should always transcend academia.

I would like to thank my committee members, Professor Deniz Zeyrek Bozşahin and Assistant Professor Murat Perit Çakır, whose professional help conveyed my study when I was stuck with interpreting the results.

In addition, a thank you to Professor Cem Bozşahin, who has special contributions to my being here today by allowing me to attend the department interviews, although I was eliminated in the preselection process, and allowing me to show my enthusiasm for the academia. Also, a thank you to Assistance Professor Hulusi Kafalıgönül, who introduced me to Cognitive Neuroscience when I took a class from him years ago, whose enthusiasm for the “underlying structures” had a lasting effect.

The members of my dissertation committee have generously given their time and expertise to better my work. I thank them for their contribution and their good-natured support.

I am also thankful to all my friends, who were continuously supporting me and gave me their help whenever I needed it, by creating an interdisciplinary point of view to my study. Lastly, a big thank to the participants of this study, all of them were so collaborative and enthusiastic about the experiment, their participation was the most important element of this dissertation.

TABLE OF CONTENTS

LIST OF TABLES

TABLES

LIST OF FIGURES

FIGURES

LIST OF ABBREVIATIONS

F1	The most familiar voice to the listener; the voice of the Participant's couple
F2	The selected most similar voice to the participant's couple from the pool
F3	The selected most dissimilar voice to the participant's couple from the pool
ADP	Auditory Distance Perception
SDT	Specific Distance Tendency
SPL	Sound Pressure Level

CHAPTER 1

INTRODUCTION

1.1 On the use of auditory cues

In the human understanding of the external world, the auditory environment is important as much as the visual environment. Likewise, in the perception of the acoustic environment, distance perception is as important as localization. For instance, when the source is out of the visual field, the perceiver is in dark room, or to eliminate the cognitive effort of visual attention, auditory processing must convey all the spatial information, including the distance information of the source. In such a case, when the visual information from spatial perception is not available, the auditory mechanism provides it. This type of acoustic processing includes using different cues such as spectral, acoustical and cognitive ones.

Some studies on auditory distance perception show that processing auditory assessment provides a scaling of sound sources and provides to focus the listener's attention on the closest sound source. As a result, people tend to overestimate the distances of near sound sources, whereas they tend to underestimate the distances of sound sources located far away [Zahorik, 2002a]; [Brungart and Scott, 2001]. Contrary to this, the estimation of auditory distance perception is accurate for sound sources positioned within approximately 1 m. It is assessed as 'mostly accurate' because of the fact that, human auditory distance perception is not perfectly accurate in any range. Based on this information, it may be stated that, as the sound source moves away from the egocentric distance, the accuracy of distance altered regarding two conditions. Those are whether the source distance overestimated, if it is in peripersonal space, for example, the sounds that are audible within reaching and grasping distance (approximately 1 m. from the listener) or the distance underestimated, if it is a distant sound in extrapersonal space [Zahorik, 1996]. Although a considerable amount of research have been carried out on the auditory distance perception, the research on the cognitive aspects has been rather limited.

It can be stated that perception of the sound source localization is an area that largely studied than the distance of the sound source. The reason is that some of the auditory cues to estimate distance and localization of sound sources are uncertain for listeners, and this uncertainty or ambiguity probably originates from the multidimensional contribution of auditory and cognitive cues on auditory distance perception, it causes auditory distance perception less accurate and limited for further studies. To clarify, the ambiguity of auditory cues, it can be assessed that, binaural cues and the masking effect of the environmental/background noise can cause a suppression to the sound source, and because of this compression the distance of a sound source can be perceived less accurately [Loomis et al., 1998]. According to Blauert's study, the judgments of distance on an acoustic field are more complex, require more cognitive effort and interaction with different acoustic cues than the judgments of sound direction, also their results are much harder to model and come up with a rigorously defined objective or subjective scale on source distance [Blauert and Butler, 1985]. In accordance with this study it can be emphasized that in order to process the acoustic information about the distance of a sound source, listeners must convey much more cognitive effort rather than the processing of direction information.

Therefore; according to recent and past studies perception of sound source distance may have been computed from a multidimensional processor with various pieces of information including both acoustic and non-acoustic factors. It may have also be aided by directional localization or even cognitive cues such as vision, familiarity, decision making or learning. Whereas directional and spectral localization of sound sources are mostly focused on cues, some researchers cast new emphasize on the basis for judging the third dimension, as known as sound source distance [Zahorik, 2002b], [Kolarik et al., 2015].

Among previous studies for auditory perception, although directional localization to detect a sound source is extensively studied, less is known about cognitive cues that allow and aid auditory distance judgment. However, among all the cues that the assist and enhance auditory distance judgment, effects of familiarity has not been identified as a rigorously defined cue for source distance perception. Any objective or subjective scale does not exist that can be reliably used to assess familiarity with a subject with a given sound source. Because of a lack of objective or subjective familiarity scale, there is a need for a carefully investigated study on the possible effects of familiarity aspect on auditory distance perception.

1.2 Background

Several recent studies indicate that different aspects of the familiarity effect as a cognitive cue to enrich auditory distance perception. For example, Wisniewski et. al. studied native language (semantically and lexically familiar) vs. foreign language (semantically unfamiliar, lexically familiar) vs. backward speech (semantically or lexically unfamiliar) to investigate how a familiar speech sound can enhance estimation of source distance. The study indicated that familiarity on speech properties could improve the accuracy of estimation both far and near sound sources [Wisniewski et al., 2012]. Moreover, Zahorik states that familiarity with the surrounding environment (i.e. its reverberation features) can enhance the accuracy of distance judgments of the sound source. [Zahorik, 1996] Sound intensity alone is insufficient for the listeners to determine the actual distances to an unfamiliar sound source since its original sound intensity is unknown to the listener[Bronkhorst and Houtgast, 1999]. However, with increasing familiarity with both a given sound and surrounding environment, the distance judgments based on sound intensity can become more accurate [**?**]. Briefly, the overall effect of familiarity as a non-acoustic cue indisputably plays an enhancing role in distance judgments of a sound source.

Although the aid effect of familiarity on auditory distance perception about the aspects of familiar speech, familiar environment or familiar acoustic cues were studied separately or all together, the effect of interpersonal familiarity has not been systematically studied before.

This thesis has an objective of studying auditory distance perception by investigating the non-acoustic cues, particularly the familiarity effect in cognitive aspects and whether a personally familiar voice can enhance the accuracy of the estimation.

In brief, there is now a considerable and a growing amount of studies indicating that auditory distance perception enables valuable spatial information that guides human behavior in different acoustic environments. Recent studies on auditory distance perception has provided valuable information about the neural processes underlying auditory distance perception, the results of sensory loss and how perceived auditory distance is adaptively biased to cope with immediate threats [Kolarik et al., 2015] On the other hand, there is still a lack of knowledge in our understanding of auditory distance processing.

1.2.1 Terminology

Commonly used terms are clarified below, for a better understanding of the context.

- **Auditory distance perception** is the general term used for estimating distances into two conditions: As it is demonstrated in the Figure 1.1, **1)** between the listener's ear and sound sources or **2)** distances between sound sources, in any and all directions to the listener. It also includes the ability to hear the distance changes (moving or looming sound sources) of sounds from near to far or/and far to near and at varying angles [Zahorik, 1996].

- The term **depth perception** defines the auditory distance of a sound source straight ahead of the listener. Hence, sensory depth perception may be described as looking straight into a corridor and estimating the distances of sources using auditory cues 1.1. Note that, this term is used for when visual cues aid to the auditory perception [Kolarik et al., 2015].

- **Auditory distance cues** enable a listener to perceive the distance of sound sources. Those cues can support the perception in a multidimensional way in order to collect more accurate data from the sound features and the environment [Mershon and Bowers, 1979].

- **Peripersonal space** is the acoustic field within reaching and grasping distance (approximately 1 m. from the person)

- **Extrapersonal space** is the acoustic field farther from the reaching distance.

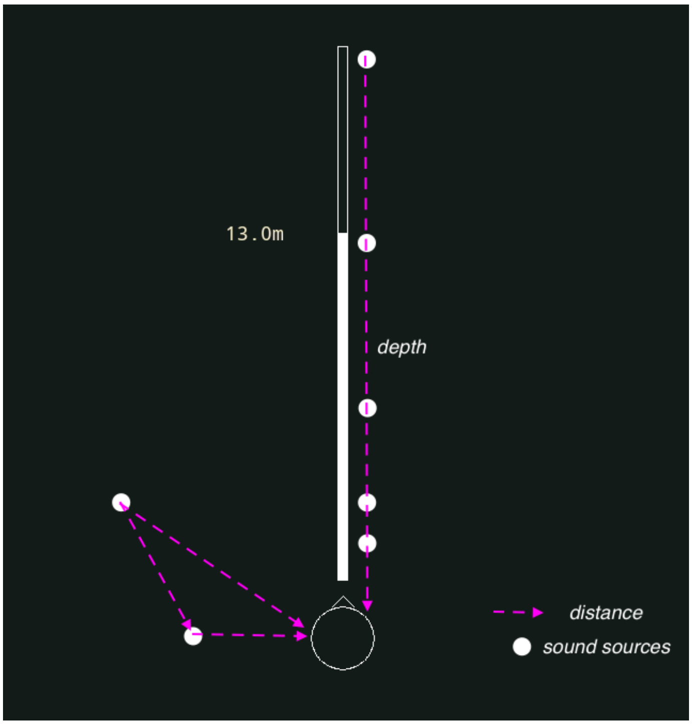

Figure 1.1: Graphical demonstration of the auditory distance and depth perception

In the experiment, one-dimensional sound field provided to the listener, regarding a line starting from the line in the middle of the ears toward the sound source. In this study visual cues were not used for determining the distance of a sound source, rather an empty room with a plain computer instructions were used. This kind of restriction of avoiding visual aid renders only the auditory cues.

1.2.2 Purpose of this Thesis

This thesis investigates three primary objectives relative to auditory perception; The first objective is contributing to the cognitive and psychoacoustics research on auditory distance perception. Even though the past studies has given more attention to auditory distance perception , some aspects still require further investigations, namely cognitive cues that aid auditory distance perception .

Secondly, the goal is to investigate the effects of cognitive cues on auditory distance perception . In particular, this study has an aim of contributing a research on the aspects that have a lack of studies on the human ability to perceive auditory distance. Although the non-acoustic cues are commonly considered in past studies, such as learning effect and visual contribution to distance perception, this thesis focuses on the familiarity effect.

Finally, the key point is to investigate the familiarity impact at the low-level contribution. To our knowledge, familiarity as a non-acoustic cue has been studied only at high levels and in a combination of the learning effect. Furthermore, this thesis investigates the familiarity cue in the level of interpersonal voice familiarity with three sub-levels on the aspect of auditory distance perception.

CHAPTER 2

AUDITORY DISTANCE PERCEPTION

Auditory distance perception is a multidimensional phenomenon which is highly relevant for daily routines, also for an avoidance of obstacles or awareness of auditory stimuli in the environment [Kolarik et al., 2015]. There are several cues that aid auditory perception. Even though, the cues that mainly contribute to perceived distance can be divided into two; acoustic and non-acoustic cues, as it is mentioned that distance perception works with a multidimensional map. This multidimensional contribution of auditory factors that convey auditory distance perception will be illustrated in this chapter. The aim is to give a broad view of current research and explain the grounds for this study.

2.1 The Human Ability to Perceive Auditory Distance

The cues that are used by the auditory system for directional perception are well-known: differences in time and level between the two ears provide robust azimuth information, while monaural spectral cues arising from filtering effects of the head, shoulder, torso, and outer ears give elevation information and partially solve spatial confusions [Blauert and Butler, 1985]. It is known that humans can localize sound sources with a high level of accuracy with less that 1° error when sources are at the front direction [Langendijk and Bronkhorst, 1999]. The human ability to perceive distance is a complex phenomenon which has a lack of accuracy in comparison with directional localization and likely to be computed from multidimensional factors. By considering the recent studies, there is a dramatically increased interest in auditory distance perception, [Rungta et al., 2017]. There are extensive reviews on previous research on distance perception indicating the growing interest on the subject. In this section, the aim is demonstrating the overall knowledge on distance perception and gathering the relevant studies to clarify the fundamentals of the study.

2.2 Auditory Cues to Distance

The perception of sound source distance is known to depend on a variety of acoustic cues, including: as monaural cues; intensity [Mershon and King, 1975], frequency spectrum [**?**], the ratio of direct-to-reverberant sound energy [Gardner, 1969] and binaural cues [Bronkhorst and Houtgast, 1999]. In addition to those cues, some cues can be categorized as non-acoustic cues that possibly effective on auditory distance perception , but very little is currently known regarding the all alone effects of non-acoustic cues especially the familiarity effect.

In other words, auditory distance information conveyed upon four primary static acoustic cues, relatively as; Intensity, Direct-to-reverberant energy ratio, Interaural Level Differences (ILDs) and Spectrum. Most of the auditory information collected from those acoustic factors [Zahorik, 2002b].

Sound Intensity factor is an important piece of information for distance perception. It can be directly assessed as, when the sound is getting far from the listener, its intensity decreases relatively [Mershon and King, 1975]. For the far field sources in free space, the intensity level varies inversely with the square of the distance. In the real world, the environment is mostly reverberant and due to the reflective surfaces in the sound field, such as walls, ceiling, trees, mountains. In such circumstances, reflected sounds will add up to the direct sound, and the intensity increases at the field location contrary to the inverse square law predict. Nevertheless, the sound intensity is still used due to its reliability as an absolute auditory cue for distance perception. At some distance, the direct sound and the reverberant sound components will have the same energy, this distance is known as *the critical distance* or *reverberation radius*. Beyond that distance, the reverberant sound dominates the direct sound and judgment of the distance by intensity level become misleading [Tervo, 2009]. As it is shown in the Figure 2.1 the direct sound and the reverberant sound components have an equilibrium point at the distance of the *reverberant radius* defined. Inside the critical distance only the direct sound dominates, however, outside the critical distance the reverberation is much more dominant. It shows that only the intensity factor is not enough to get adequate distance information from a sound source in real life.

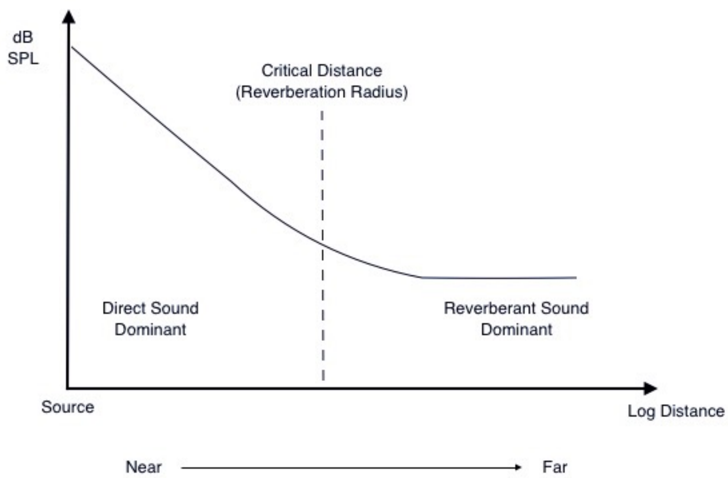

Figure 2.1: Graphical demonstration of the critical distance

Secondly, one of the auditory factors that contribute auditory distance perception

Direct-to-reverberant energy ratio is known to be another important acoustic cue to sound source distance. It decreases as the distance between sound increases of the source and listener, since the intensity of the direct sound changes with the inverse square law while the reverberation remains approximately the same. There is $1/R$ interaction between direct, first arriving, sound and the reverberant sound [Blauert and Butler, 1985]. The sound intensity goes in a straight line from the source to the listener's ear, and its level decreases by 6 dB for logarithmically increased the source distance [Kolarik et al., 2015]. The reverberant sound energy is reflected from surfaces, such as walls or objects, before reaching the listener' ear. It can be estimated by a distributed sound field with constant energy notwithstanding of the sound source location if the room is wide; the level of the reverberation of sound alters only inconsiderably with distance [Zahorik, 2002b].

The other auditory factor is **Spectrum**; for very distant sound sources approximately more than 15m, sound-absorbing properties of the air slightly attenuate the perceived level of high frequencies [Blauert and Butler, 1985]. Briefly, as distance increases, especially for distances more than 15m, the amount of high-frequency attenuation increases [Zahorik, 1996].

The last factor in auditory distance perception can be classified under the name of **Binaural cues**. Interaural cues depend on source distance, and this is known as the auditory parallax [Zahorik, 1996]. For nearby sources, ITD and ILD cues are more prominent and vice versa for distant sources. In other words; when the sound sources were located in the near-field, intensity level and time differences will be no longer independent of radial distance, while they are dependent for far-field planar waves. These differences in both intensity and time often associated with differences resulting from acoustic parallax, as mentioned above. Sound source distance changes which located in near-field also provide alterations in the spectrum reaching the listener's ear, due to diffraction around the head and pinna as identified by head-related transfer function (HRTF) measurements [Brungart and Rabinowitz, 1999].

2.2.0.1 Familiarity as Auditory Cue to Distance

One of the non-acoustic cues that contribute auditory distance perception is stimulus familiarity is important as acoustic cues that mentioned above. This cue also knows as the learning effect. Some studies are showing that repeated exposure to an unfamiliar sound stimulus increases the accuracy of auditory distance perception judgments. Coleman Colman1962FailureSound assessed in his study that participants' accuracy enhanced over successive trials when judging distance to noise bursts presented in an acoustic free field (for distances between 2.7 m and 8.2 m). Mershon et al. Mershon1979AbsoluteDistance also reported that judgments of distance for noise bursts became more accurate over a repeated sets of trials in a reverberant room (for distances between 0.75 m and 6 m). Shinn-Cunningham Shinn-Cunningham2000DistanceSpace presented a similar finding: that participants presented with noise bursts up to 1 meter away (outside of the peripersonal space) provided distance judgments with improving accuracy over many trial sets performed in a period of 3-5 days. It is reported in a study by Brungart and Scott that listeners rely on prior, long-term knowledge of speech intensity rather than the experimental context of listening experience. When judging the distance of a sound source, whispered speech always judged to be nearer and shouted speech is further due to the prior experience of intensity levels of a speech Brungart2001TheSpeech..

Familiarity was also investigated in the context of language familiarity. For instance, Wisniewski et al. compared distance judgments of native English speakers to different language stimuli: English, Bengali and time-reversed English and Bengali speech. It is reported that participants judged the distance of normal speech more accurately than that of time-reversed speech and accuracy did not differ between English and Bengali speech, those results assessing that distance discrimination of speech sounds based on phonetic rather than lexical familiarity.Wisniewski2012FamiliarityAcuity

2.3 Models of Auditory Distance Perception

2.3.1 Bronkhorst and Houtgast's model

As one of the models in auditory distance perception, Bronkhorst and Houtgast proposed a computational model in order to estimate human ability to auditory distance judgment in a controlled condition where the direct-to-reverberant-energy-ratio (DRR) cue is dominant, demonstrated in Figure 2.1. The model indicated an accurate estimation of listeners' perceived distance depending on the prior knowledge of certain acoustic attributions of the environment by using monaural data. This model is successful in predicting the apparent distance results reported in related studies. Other models based on binaural cues used either prior knowledge of environment (e.g. room impulse responses [Bronkhorst and Houtgast, 1999]) or extensive training data (e.g. learning effect [Vesa, 2007]) to judge sound source distance. While these studies demonstrate that distance judgment can be further improved with binaural input, primarily based on DRR cues.

In the model, the time window of 6ms used for the purpose of include early reverberant sound source from the nearby surfaces. The contribution of these early reverberant sound source to the direct sound energy increases with distance until the reverberant sound source is dominant. The auditory horizon takes place at that point, where the reverberant sound source is dominant because, the estimated direct sound energy will hardly change when the distance is increased further, as a consequence of the reverberant sounds.The direct sound energy is considered to lie within the first 2.5 ms of the sound representation; this time-window is where the direct sound source energy is dominant to the DRR [Brungart and Rabinowitz, 1999]. According to the model that Bronkhorst and Houtgast proposed, the human auditory mechanism can use the DRR cues, to estimate distance. This model has shown to converge with other research results used by Zahorik in its power function fit analysis. The amount of reverberation is assessed by measuring the ratio between the direct sound energy and the reverberant sound energy.

2.3.2 Zahorik's model

Zahorik assessed as a result of his studies that the ability of humans to judge the distance to a source of sound is much less accurate than the capacity to determine the angular direction of a sound source. This inaccuracy on auditory distance judgment results with a tendency of overestimating the near sounds, especially within the range of peripersonal space (approx. 1m), they perceived as if it was coming from farther than its real distance while they perceived far distances as nearer than its real distance by underestimating it. It can be concluded that perceived auditory field is a distorted or biased representation of the physical, acoustic field [Zahorik, 2005]. It is essential to characterize the association between perceived distance and source distance in the form of a psychophysical function, to investigate the issues related to distance localization bias.

Zahorik proposed a model know as power function fit analysis to estimate human auditory distance judgment distortion. This model includes applying Steven's Power Law/ Least-Squares Fitting to the distance responses. The logic behind the function is; equal stimulus ratios produce equal senses ratios. Stimuli; in this case the real distance of the sound which appears to stand in a particular ratio, stimuli are provided by stimulus energies which are in an another specific ratio.

$$S = k.R^n$$

That is, psychological intensity, known as Sensation in humans, increases as the n-th power of stimulus intensity; when k is simply a scaling constant. When the logarithms of both the Real distance (R) and the listener's judgment values (S) are taken, the judgments should fall in a straight line, whose slope shows the amount of compression. This compression function indicates each subject's distortion of auditory distance judgment depending on the ranging distances of sound sources.

CHAPTER 3

MATERIALS AND METHODS

Given these points about the background of auditory perception, this study was conducted due to the lack of research in voice familiarity aspect and to convey a particular study. There were two separate parts of the experiment that the participants had to attend. In the first section of the experiment, appointments with twelve couples in different time slots arranged at the METU SPARG (Spatial Audio Research Group) Lab, which is located at the Modeling and Simulation building in METU. All couples were strangers to other participated couples. Their speaking voices separately recorded in the silent room (35 dB). All participants instructed to read the five sentence in their most neutral and normal way of speaking tone.

3.1 Part I: Recording speech samples

3.1.1 Material

3.1.1.1 Ethics Statement

The study was approved, over a protocol code: *2016-FEN-055*, by the Applied Ethics Research Center, Middle East Technical University, Ankara. All participants were asked to read and sign an informed consent form before participating in the study.

3.1.1.2 Room

The silent room where the Experiment Part I and II were conducted has eleven squaremeter ($11m^2$) floor area. Acoustical parameters of the room were T30 $\cong$ 80 ms.

3.1.1.3 Talkers

A total of twenty-four volunteered participants that consist of twelve heterosexual couples, ranging in age from 19 to 30, invited to participate in the study. They were together with their partners at least more than a year with daily interaction. Also, all the participants reported that they do not have any hearing-related problem.

3.1.1.4 Sentences

Five different Turkish utterances which are constituted from the words that had been scaled as emotionally neutral in Turkish language [Çağlayan and Özkurt, 2017], all sentences have equal number of syllables (6 syllables) recorded at each production level for each of the twenty-four talkers, they have been told to avoid any emotional prosody while they reading the utterances;

1. Dünya Gezegendir.(EN: The earth is a planet.)

2. Benden kazak aldı. (EN: S/he took a jumper from me.)

3. Ormanda ağaç var. (EN: There are trees in the forest.)

4. O, radyoyu açtı. (EN: S/he turned the radio on.)

5. Kaşık çekmecede. (EN: The spoon is in the drawer.)

3.1.2 Procedure

Before each set of measurements, the microphone adjusted to the height of the talker's mouth, and the microphone placed at a 1m distance from the talker's chin. Then the speakers were instructed to begin speaking in their most neutral and normal speaking tone. Figure 3.4 is showing the recording settlement. Their speech recorded at a sampling rate of 44,100 Hz, using a microphone.

Figure 3.1: Recording Settlement

3.1.2.1 Processing the data

Altogether, for each stimulus presentation, familiarity conditions were created by using gammatone filterbank and measuring the similarities between sound's power spectra at each of the frequencies using the correlation algorithm (AppendixB.1) It is accepted that gammatone filters can simulate the performance of the human auditory peripheral mechanism. Primarily, it is used in dealing with the robustness of speech recognition systems [Lewicki, 2002]. In this study, it is used as filtering the speech recordings as much as the human ear does, and after that sounds compared with each other in the scope of their power spectra levels. By this way, the most and least similar audio files can be detected for each speech recording. All audio files were divided by their corresponding familiarity levels, sentence numbers and convolved with binaural room impulse responses (BRIRs) measured as described in the following section.

Objective familiarity defined as the distance between the auditory magnitude spectra. All listeners sound pool composed of three different categories. These categories are called F1, F2, F3. The category Familiar (F1) consists of utterances belong to listeners real-world partners. The category F2 and F3 constructed according to spectral resemblances to listener's real-world partner. The sound files in category Near-familiar (F2) were chosen as the most similar and the category Unfamiliar (F3) were selected as the most dissimilar. All the categories include convolved sounds coming from six different distances.

3.1.3 Processing of speech signals

An observational study of auditory distance perception is usually not feasible. A practical approach involves presenting the subjects with binaural recordings which simulate natural spatial hearing scenarios with audio presentations over a pair of headphones. This is also the method used in this thesis.

3.1.3.1 Measurements of binaural room impulse responses (BRIRs)

The hallway which measured $18m \times 1.8m$ was used for recording the binaural room impulse responses with the dummy head microphone, Neumann KU-100 (AppendixA.1. The microphone was set at the height of 1.5m and placed at one end of the hallway. The distance to the right and the left wall was set at 0.9m, as the middle point. By placing the speaker at certain distances that were used in the experiment (0.5, 1, 2, 4, 8, 16 m), These binaural room impulse responses (BRIRs) were collected from relative distances.

The experimental task consisted of 90 speech samples from F1, F2 and F3 conditions convolved with the binaural room impulse responses obtained via recordings made in a hallway with the KU-100 binaural microphone. The primary experimental variable in each stimulus presentation was the simulated distance of the talker. A total of six logarithmically ranged simulated distances were used in the experiment: 0.5, 1, 2, 4, 8 and 16m as convolving parameters. A binaural impulse response has been measured with each distance in a hallway using the logarithmic sine sweep method [Farina, 2007] and convolved with dry speech samples recorded in the lab. and convolved with previously recorded speeches. Spectral and temporal cues given by distance, elevation, and azimuth changes of a sound source to a listener are captured by KU-100 binaural microphone [Gardner, 1969] There are six, (azimuth 0 °) distance measurements. All of the recordings as mentioned above were captured at a sampling rate of 44.1 kHz in the front direction.

3.1.3.2 Sound Pressure Level Calibration

Since distance perception relies heavily on the sound intensity and this would confound the results of a study investigating familiarity, each speech stimulus has to be equalized for level prior to the subjective experiment. This equalization was carried out in the following way:

1. A 1 kHz pure tone is played back from the headphones (Superlux HD-330) and the reproduction level was measured using a miniature microphone positioned at the entrance of the right ear canal of the dummy head microphone using a sound level meter. The level of the microphone amplifier was fixed when a reading of 80 dB SPL was obtained.

2. Speech signals processed with BRIRs were normalized for their total energy.

3. The processed speech signals are then played back via the same pair of headphones and the sound level was measured. The average sound levels (L_p and L_{eq}) are given below:

Table3.1: The values of the Instantaneous Sound Pressures measured over a specified period of time, measured in decibels (dB).

SPL Measurements	**FEMALE Voices**	**MALE Voices**
Peak Level	77.27 dB	77.47 dB
Equivalent Level	66.12 dB	65.91 dB

It may be seen that these values are generally in agreement with sound levels of normal speech when spoken from a distance of about 1 m.

3.2 Part II: Measurement of Auditory Distance Judgments

3.2.1 Material

In Part II, the experiment was conducted on a Macintosh Mac book Air. The Open Sesame software was used to set up and control the experiment [Mathôt et al., 2012].

In this part, appointments with eleven couples, whose voices recorded in Experiment Part I, in different time slots were arranged at the same location before. All participants attended the experiment alone by using headphone and sitting in front of a computer in the METU SPARG Lab. The room was darkened except for the light from the computer screen to alleviate the effects of visual feedback confounding the experimental results.

3.2.1.1 Listeners

A total of twenty-two volunteered listeners (eleven male, eleven female) were used in this part. The number of the participants who had participated in the Part I was twenty-four (twelve male, twelve female) volunteers.

All had reported that they have normal hearing and their ages ranged from 21 to 30. None of them had previously participated in distance perception or any other kind of psychoacoustic experiments.

After the experiment, one participant indicated that he did not fully understand the experimental setup. For this reason, the data of the participant has discarded from the experimental data.

One of the couples did not attend as listeners in the experiment part II. However, all voices of the couples who participated in the experiment part I, they were used as stimuli for the listeners who participated in part II. In other words, for each participant a personal sound file pool created from all the recorded sounds. The personal sound file pool included the participant's partner's voice recordings along with its distance convolutions and the least and most similar voice recordings to the source voice (F1). Listeners, who participated in Part II, only listened to voices from their own pool of audio files.

Lastly, in the stage of the preprocessing of the data, five participants were excluded as outliers from the experiment data on the grounds that they could not fully understand the experimental setup or gave random responses with a very high variability. This elimination process of those subjects will be discussed in the **Results and Discussion** chapter.

3.2.2 Experimental Procedure

The recorded speech stimuli that were used in each experiment were presented to listeners who were seated at the center of the room, as it is shown in the Figure 3.2. In the experiment, lights were off, and the door was closed.

While the presentation level very slightly varied for each stimulus, the variance of the level was less than about 2 dB for most cases. The stimulus presented to the participants were randomized and each stimulus were presented once.

Figure 3.2: Experiment Part II settlement.

For the experiment part II, the listeners were positioned facing the computer, and they were asked to make judgments about the speech stimulus after they have listened to it from the headphone. The slider screen appears after the corresponding audio file played, and the desired distance is selected between 0.0m and 20.0m with the touch pad by the listener. The Figure 3.3 is showing the slider screen.

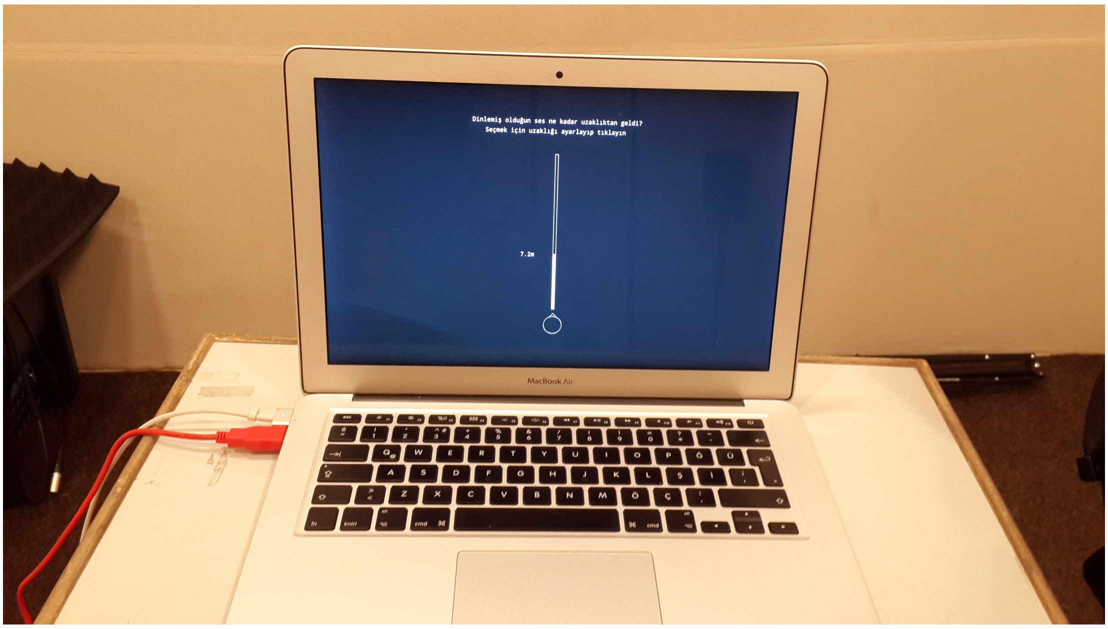

Figure 3.3: Slider Screen of the Experiment

After first instructions, listeners started a training session to get them acquainted with the user interface as well as the experiment. In the training task, practice sequences (playing speeches (different from the experimental sequence) three sentences from the nearest (50cm) and three sentences from the furthest (16m) distances without giving any feedback) and the experimental task that consists of ninety convolved speech samples from F1, F2, and F3 conditions. The practice trials were inserted to expose the participants to the distances of presentation of the speech stimulus.

To investigate the effect of speaker familiarity on the results, three different levels of familiarity were tested. For the first condition of familiarity, the recorded speech of the listener's partner was selected for the listener's audio file pool. The second condition was conducted by selecting the audio files of another participant who had the vocal characteristics most similar to the partner's voice. The latter condition of the familiarity was created by selecting the most dissimilar voice to the partner from the pool. Thus, they have listened to ninety audio files with three factors. Those factors were as following;

- Three levels of Familiarity

  -The Familiar Voice denoted as F1 (Partner's voice)

  -The most Similar Voice, F2 (Selected by using gammatone filtering, see Section 3.1.2.1)

  -The most Dissimilar Voice, F3 (Selected by using gammatone filtering, see Section 3.1.2.1)

- Five Different Sentences: selected from Caglayan's study from the sentences that scaled as emotionally neutral utterances [Çağlayan and Özkurt, 2017].

  1. Dünya Gezegendir.(EN: The earth is a planet.)
  2. Benden kazak aldı. (EN: S/he took a jumper from me.)
  3. Ormanda ağaç var. (EN: There are trees in the forest.)
  4. O, radyoyu açtı. (EN: S/he turned the radio on.)
  5. Kaşık çekmecede. (EN: The spoon is in the drawer.)

- Six different distances: The decision to select these particular distances was based on past reviews of auditory distance perception. The distances are ranged logarithmically as Brungart and Scott used in their auditory distance perception experiment [Brungart and Scott, 2001]. This way, a variety of distances, both close and far away from the measurement position were obtained.

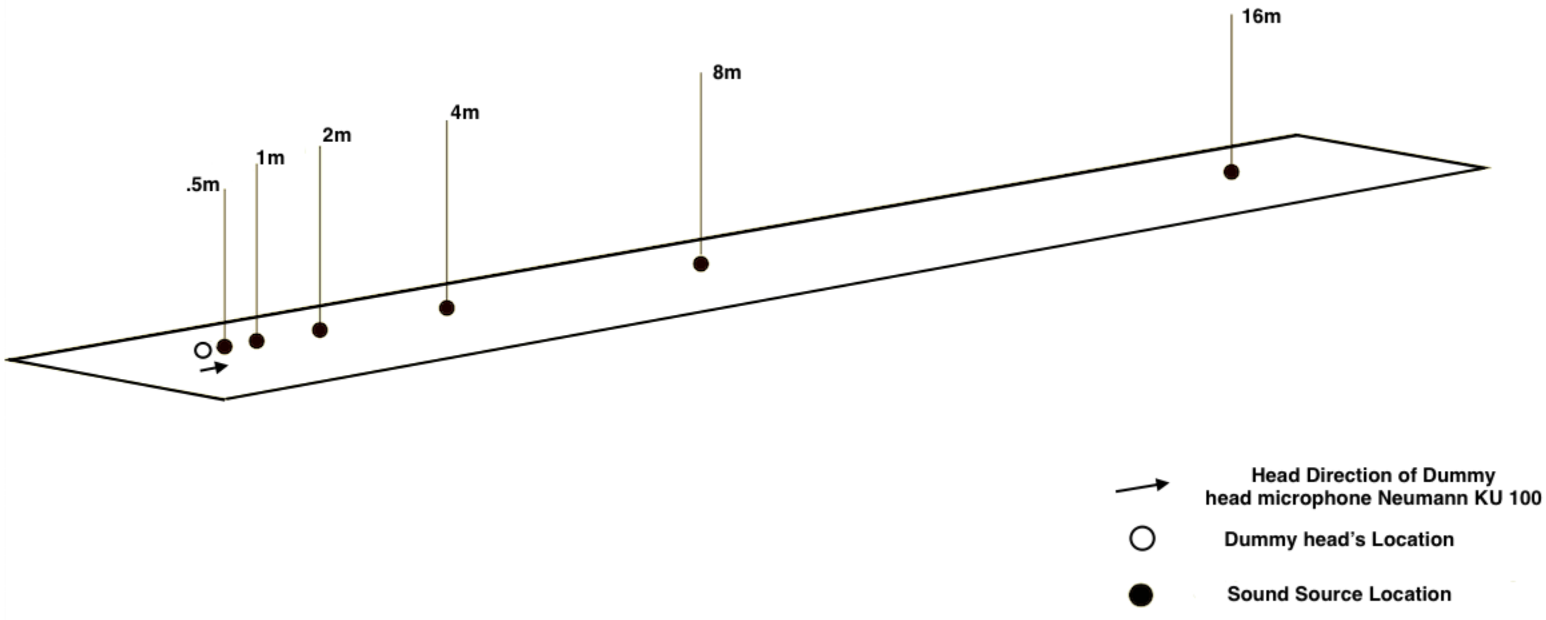

Figure 3.4: Demonstration of the Real Distances of sound sources presented in the experiment

CHAPTER 4

RESULTS AND DISCUSSION

4.1 Analyzing Data

There were two levels of analyzing the data obtained from the listeners. Firstly, the following method was used as a replication of a previous auditory distance perception study [Zahorik, 2002a]; the power law function was fitted (least-squares criterion) to the geometric means in each listener's judgments. Stevens' power law provides a general mapping between actual and the perceived intensity of a stimulus [Zwislocki, 2009]. Zahorik (2002b) developed a similar model as explained in Chapter 2. A similar approach is used here to map the actual distance of the stimulus to its perceived distance. The mapping between the actual and perceived distance is given as;

$$\bar{D} = \beta . D^{\alpha}$$

Where y is perceived distance, β is constant, α is the power law exponent and x is the given and fixed source distance value. The α in the function indicates the amount of non-linear compression where $\alpha < 1$. The obtained data was first transformed and fitted using SPSS with this curve model and the parameters are calculated. All judgments of sound source distance were plotted (log-log axes) below for each listener, and within those plots, point symbols indicate all the responses on individual trials for the participant. Also, the square symbols represent geometric means of responses for each distance.

Linear fits on logarithmic coordinates to the geometric means are represented in each plot, together with their corresponding a values (compression value). In all cases listeners judged distances to be externalized from their head [1]; thus there was no response for the 0m. On the other hand, the proportion of variance imputed by curve fitting is low (R^2ranging from 0.1 to 0.44), this may be caused of listeners did not participate in any other auditory perception experiments before. A very low α corresponds to a flat response indicating either that i) the subject responds completely randomly, or ii) responds consistently the same distance regardless of the stimulus. Therefore, subjects with an α that is one standard deviation below the mean are eliminated. For this reason, listeners whose alpha value below the one standard deviation (S.D: 0.13) were

[1] Inside the head localization [Plenge, 1974] is a phenomenon which may be observed with binaural audio presentation where the subjects can tell the direction of a sound source but perceive the source to be inside their craniums close to their auditory egocentric position.

omitted from the data, considered as they did not fully understand the experimental setup or were assumed to respond randomly to the distance parameters. For example Subjects 16 and 17 in 4.1 are two such examples.

In Figure 4.1, dots show raw distance judgments $\bar{D}$: 15 replications/distance. Square symbols indicate geometric means fitted to the data. The geometric mean is appropriate while handling a skewed data. The use of a geometric mean "normalizes" the ranges being averaged so that no range dominates the weighting, and a given percentage change in any of the properties has the same effect on the geometric mean. Data from each condition were fit with a power function of the form $\bar{D} = \beta.D^{\alpha}$. Where $\bar{D}$ is perceived distance, β is the constant, α represents the power-law exponent, and D is the given source distance. Fit parameters are shown on each panel. For details see Appendix C

After eliminating the problematic listeners, the average value of a was calculated to be approximately 0.32 across all listeners, which is substantially less than 1.0 (which indicates a perfect linear fit). The mean value of k was approximately 1.1, slightly higher than the veridical $k > 1$. These results show similarity with Zahorik's model [Zahorik, 2002a], [Anderson and Zahorik, 2014], which indicates that listeners tend to underestimate source distance for most sources and that the amount of underestimation depends on exponentially actual source distance. In other words, for nearby sound sources, especially within the range of peripersonal space (approx. 1m), people tend to perceive the distance as if it was coming from further than its real distance while they perceived sources at far away as being closer. This may be related to a specific distance tendency (SDT) as assessed in [Gogel and Tietz, 1973] also, [Mershon and King, 1975].

It suggested that SDT can be applied to auditory distance perception, it can be easily demonstrated that our data have shown this tendency across all the listeners. In brief, results were coherent with other articles indicates that at far distances, the listeners begin to underestimate the real distance, and the degree of underestimation increases as the perceive distance increases approaching an asymptotic ceiling or in other words the critical distance, as it is shown in Figure 2.1 [Gogel and Tietz, 1973]; [Loomis et al., 1998].

In Figure 4.1, dots show raw distance judgments for Familiarity conditions: Where $\bar{D}$: 30 replications/distance. Square symbols indicate geometric means that fitted to the data. Data from each condition were fit with a power function of the form $\bar{D} = \beta.D^{\alpha}$. Where $\bar{D}$ is perceived distance, β is the constant, α represents the power-law exponent, and D is the given source distance. Fit parameters are shown on the each panel.

While the alpha value of F2 is a bit lower than the other two conditions, F1 and F3 have almost similar alpha values (approximately 0.34). Indicating that, in the complete data, participant's judgments to F2 (Near-familiar voice) were more variant than the other familiarity conditions. After eliminating the problematic participants from data, alpha values for the familiarity conditions were respectively as; 0.51, 0.34, 0.38 (see table 4.8). Judgments for the F1 (Familiar) condition show a more accurate relationship between target distance and estimated distance.

In Figure 4.3, dots show raw distance judgments for the five different sentences. Where $\bar{D}$: 18 replications/distance. Square symbols indicate geometric means that fitted to the data. Data from each condition were fit with a power function of the form

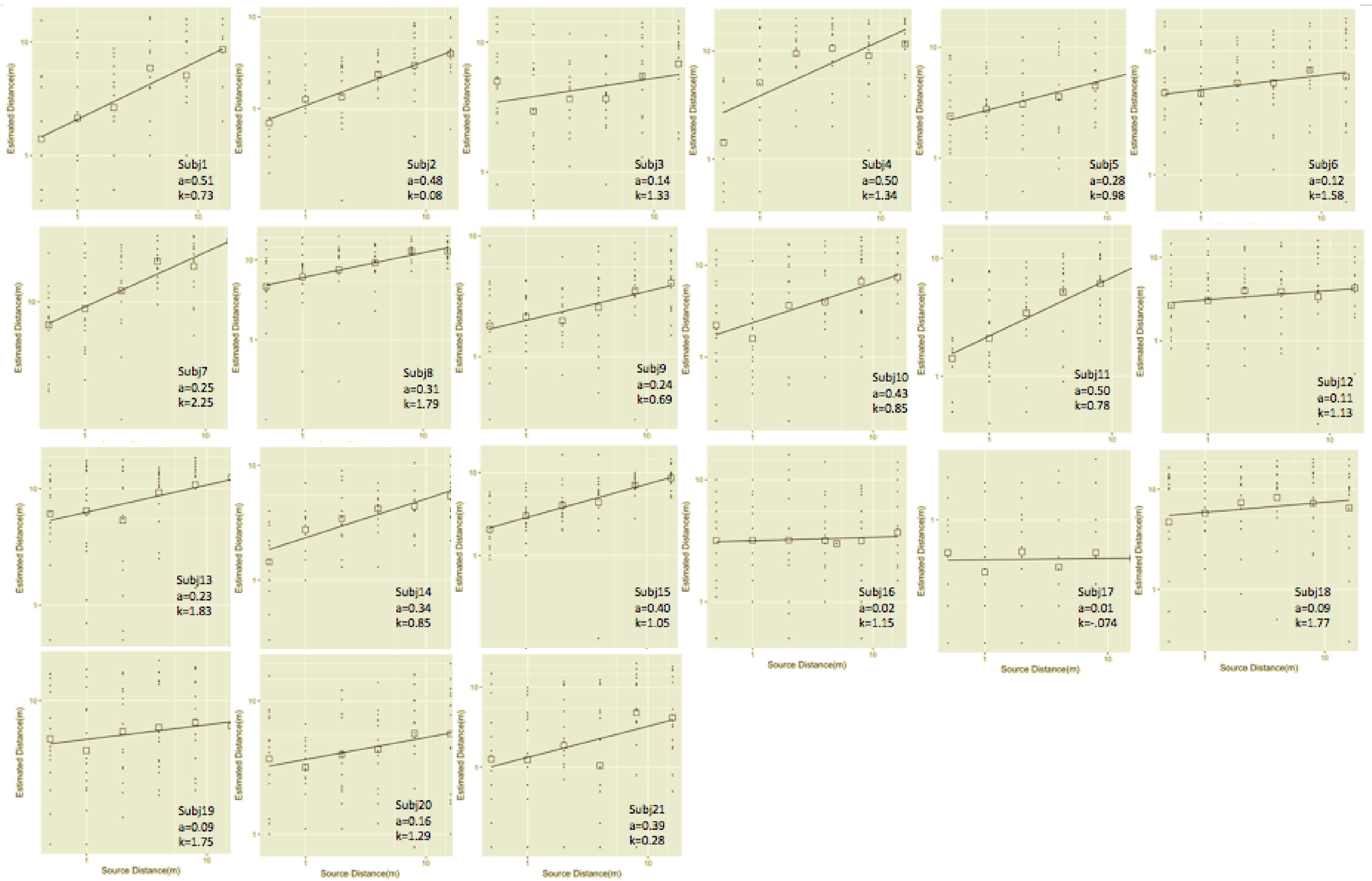

Figure 4.1: Power fits of Participants

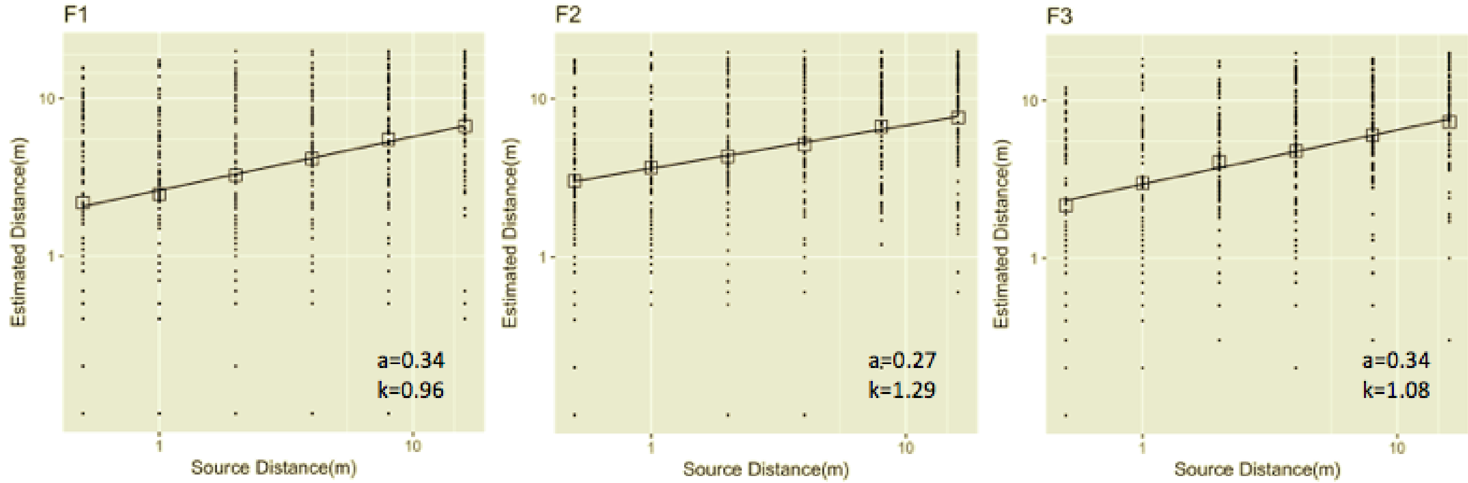

Figure 4.2: Power fits of Three Levels of Familiarity

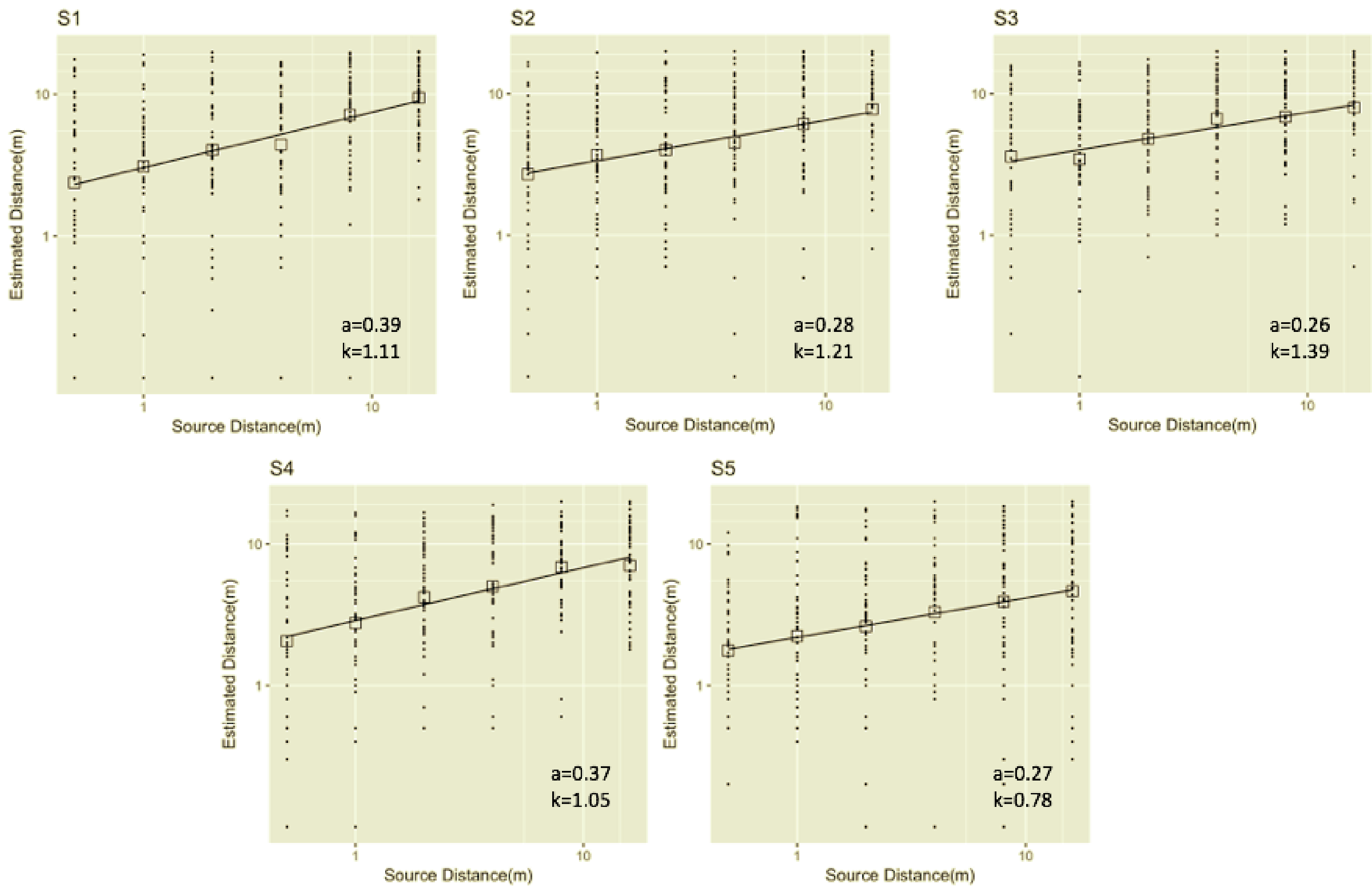

Figure 4.3: Power fits of Sentences

$\bar{D} = \beta.D^{\alpha}$. Where $\bar{D}$ is perceived distance, β is the constant, α represents the power-law exponent, and D is the given source distance. Fit parameters are shown on the each panel.

After eliminating incompatible data which have very low-level of compression, as a second step of analyzing data, repeated measures (within-subjects) analyses of variance (RM ANOVA) were used to examine the effects of the factors. The factors are based on three different levels of familiarity, five different sentences, and six different distances. The model was created as 3x5x6 factorial RM ANOVA. Since ANOVA is known to be robust with respect to the violation of normality assumption, in accordance with the papers of [Pearson, 1931] and [Edgell and Noon, 1984], the fact that the sample is non-normally distributed does not likely affect the outcome of the analysis. From this notion, the results of the RM ANOVA of this study was represented and discussed without the consideration of violation of normality assumption.

4.2 Results

Distance judgment results for all listeners are shown in the output tables were presented below. A repeated measures ANOVA was conducted for the effect of Familiarity (IV) on Distance Perception Judgments (DV) of different and emotionally neutral sentences. Familiarity included three levels (Couple's voice (F1), Similar (F2) and Dissimilar (F3) voice), the Distance condition consisted of six levels (.5m, 1m, 2m, 4m, 8m, 16m) and lastly, the Sentence condition consisted of 5 different emotionally neutral sentences. All effects were statistically significant at the .05 significance level except for the Familiarity factor.

The results of the multivariate test are shown in Table 4.1 The p-values of Distance and Sentence parameters have shown significant effects on auditory distance perception Responses, in the scope of a within-subjects effect reflected by the repeated measures. All four multivariate tests also have shown significant effects, meaning that the chosen sentences and the distance stimuli had significant effects on listeners' performance of judgment. The effect of Familiarity is marginally insignificant.

Table4.1: Output of Multivariate Test Statistics

Multivariate Tests a.

Effect		Value	F	Hypothesis df	Error df	Sig.
Distance	Pillai's Trace	,896	20.590b	5,000	12,000	,000
	Wilks' Lambda	,104	20.590b	5,000	12,000	,000
	Hotelling's Trace	8,579	20.590b	5,000	12,000	,000
	Roy's Largest Root	8,579	20.590b	5,000	12,000	,000
Sentence	Pillai's Trace	,758	10.203b	4,000	13,000	,001
	Wilks' Lambda	,242	10.203b	4,000	13,000	,001
	Hotelling's Trace	3,139	10.203b	4,000	13,000	,001
	Roy's Largest Root	3,139	10.203b	4,000	13,000	,001
Familiarity	Pillai's Trace	,289	3.044b	2,000	15,000	,078
	Wilks' Lambda	,711	3.044b	2,000	15,000	,078
	Hotelling's Trace	,406	3.044b	2,000	15,000	,078
	Roy's Largest Root	,406	3.044b	2,000	15,000	,078
Distance * Familiarity	Pillai's Trace	,305	.307b	10,000	7,000	,955
	Wilks' Lambda	,695	.307b	10,000	7,000	,955
	Hotelling's Trace	,438	.307b	10,000	7,000	,955
	Roy's Largest Root	,438	.307b	10,000	7,000	,955
Sentence * Familiarity	Pillai's Trace	,427	.840b	8,000	9,000	,592
	Wilks' Lambda	,573	.840b	8,000	9,000	,592
	Hotelling's Trace	,746	.840b	8,000	9,000	,592
	Roy's Largest Root	,746	.840b	8,000	9,000	,592

a. Design: Intercept Within Subjects Design: Distance + Sentence + Familiarity + Distance * Sentence + Distance * Familiarity + Sentence * Familiarity + Distance * Sentence * Familiarity
b. Exact statistic

Table 4.2 shows Mauchly's test for the data which is used to test for the equivalence of the variance due to different factors; the significance value for the factor Distance ($p < .001$) is less than the critical value of .05, which means that the assumption of sphericity has been violated; $\chi^2(14) = 57.94, p < .001$, therefore degrees of freedom were corrected using Greenhouse–Geisser estimates of sphericity ($\epsilon = .42$). Adjusted F values are shown in Table 4.3. While Familiarity shows no significant effect on auditory distance perception , $F(2, 32) = 2.18, p = .13$ the results of the auditory distance perception judgments were significantly affected by the distance of sound sources, $F(2.12, 33.9) = 37.9, p < .001, \eta_p^2 = .7$. There was also a significant main effect of the Sentence on judgments of the auditory distance perception judgments, $F(4, 64) = 6.9, p = .001$.

Results indicate that Sentence condition somehow has effects on people's judgments. As it is shown in Table 4.6 the 5th sentence was significantly different from the others, except for sentence 4. In order to eliminate the interaction between distance and sentence, contrasts were performed by comparing all sentences to the fifth sentence (Kaşık çekmecede / The spoon is in the drawer) (see Table 4.4). On the other hand, there were no significant interactions when comparing the condition distance to the sentences, F(20, 320) = 1,53, p = .068.

Table4.2: Output of Mauchly's Test of Sphericity

Mauchly's Test of Sphericity a.

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon b.		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
Distance	,016	57,939	14	,000	,424	,490	,200
Sentence	,318	16,505	9	,059	,681	,835	,250
Familiarity	,706	5,217	2	,074	,773	,840	,500
Distance * Sentence	0,000	.	209	.	,352	,657	,050
Distance * Familiarity	,007	61,304	54	,296	,542	,853	,100
Sentence * Familiarity	,036	43,571	35	,177	,629	,952	,125
Distance * Sentence * Familiarity	0,000	.	819	.	,225	,540	,025

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. Design: Intercept Within Subjects Design: Distance + Sentence + Familiarity + Distance * Sentence + Distance * Familiarity + Sentence * Familiarity + Distance * Sentence * Familiarity

b. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

Table4.3: Output of Tests of Within-Subjects Effects

Tests of Within-Subjects Effects						
Source		**Type III Sum of Squares**	**df**	**Mean Square**	**F**	**Sig.**
Distance	Sphericity Assumed	4898,152	5	979,630	37,903	,000
	Greenhouse-Geisser	4898,152	2,119	2311,555	37,903	,000
	Huynh-Feldt	4898,152	2,451	1998,408	37,903	,000
	Lower-bound	4898,152	1,000	4898,152	37,903	,000
Error(Distance)	Sphericity Assumed	2067,667	80	25,846		
	Greenhouse-Geisser	2067,667	33,904	60,986		
	Huynh-Feldt	2067,667	39,216	52,724		
	Lower-bound	2067,667	16,000	129,229		
Sentence	Sphericity Assumed	1092,793	4	273,198	6,904	,000
	Greenhouse-Geisser	1092,793	2,725	400,971	6,904	,001
	Huynh-Feldt	1092,793	3,340	327,228	6,904	,000
	Lower-bound	1092,793	1,000	1092,793	6,904	,018
Error(Sentence)	Sphericity Assumed	2532,535	64	39,571		
	Greenhouse-Geisser	2532,535	43,606	58,078		
	Huynh-Feldt	2532,535	53,433	47,397		
	Lower-bound	2532,535	16,000	158,283		
Familiarity	Sphericity Assumed	259,579	2	129,789	2,177	,130
	Greenhouse-Geisser	259,579	1,546	167,919	2,177	,144
	Huynh-Feldt	259,579	1,680	154,532	2,177	,140
	Lower-bound	259,579	1,000	259,579	2,177	,159
Error(Familiarity)	Sphericity Assumed	1907,534	32	59,610		
	Greenhouse-Geisser	1907,534	24,734	77,123		
	Huynh-Feldt	1907,534	26,876	70,975		
	Lower-bound	1907,534	16,000	119,221		
Distance * Sentence	Sphericity Assumed	335,710	20	16,786	1,534	,068
	Greenhouse-Geisser	335,710	7,040	47,686	1,534	,162
	Huynh-Feldt	335,710	13,134	25,560	1,534	,106
	Lower-bound	335,710	1,000	335,710	1,534	,233
Error(Distance*Sentence)	Sphericity Assumed	3500,804	320	10,940		
	Greenhouse-Geisser	3500,804	112,640	31,079		
	Huynh-Feldt	3500,804	210,144	16,659		
	Lower-bound	3500,804	16,000	218,800		
Distance * Familiarity	Sphericity Assumed	29,615	10	2,961	,254	,990
	Greenhouse-Geisser	29,615	5,422	5,462	,254	,946
	Huynh-Feldt	29,615	8,525	3,474	,254	,983
	Lower-bound	29,615	1,000	29,615	,254	,621
Error(Distance*Familiarity)	Sphericity Assumed	1868,985	160	11,681		
	Greenhouse-Geisser	1868,985	86,756	21,543		
	Huynh-Feldt	1868,985	136,405	13,702		
	Lower-bound	1868,985	16,000	116,812		
Sentence * Familiarity	Sphericity Assumed	211,908	8	26,489	1,002	,438
	Greenhouse-Geisser	211,908	5,033	42,105	1,002	,423
	Huynh-Feldt	211,908	7,619	27,814	1,002	,437
	Lower-bound	211,908	1,000	211,908	1,002	,332
Error(Sentence*Familiarity)	Sphericity Assumed	3385,002	128	26,445		
	Greenhouse-Geisser	3385,002	80,525	42,037		
	Huynh-Feldt	3385,002	121,901	27,768		
	Lower-bound	3385,002	16,000	211,563		
Distance * Sentence * Familiarity	Sphericity Assumed	449,288	40	11,232	,912	,628
	Greenhouse-Geisser	449,288	9,007	49,881	,912	,517
	Huynh-Feldt	449,288	21,611	20,789	,912	,578
	Lower-bound	449,288	1,000	449,288	,912	,354
Error(Distance*Sentence* Familiarity)	Sphericity Assumed	7881,855	640	12,315		
	Greenhouse-Geisser	7881,855	144,116	54,691		
	Huynh-Feldt	7881,855	345,782	22,794		
	Lower-bound	7881,855	16,000	492,616		

For further investigation; according to the table of Contrasts 4.4, although Familiarity had not a statistically significant effect on auditory distance perception , $(F(2,32) = 2.18, p = .130, \eta_p^2 = .12)$, contrasts revealed that Similar Voice (F2) were significantly higher from Dissimilar Voice (F3), $(F(1,16) = 5.57, p = .031, \eta_p^2 = .26)$. The other contrast for Familiarity revealed did not reveal any statistically significant effect when comparing Couple's Voice (F1) to Dissimilar Voice (F3), $(F(1,16) = .105, p = .75, \eta_p^2 = .007)$.

Table4.4: Output of Test of Within-Subjects Contrasts

Tests of Within-Subjects Contrasts							
Source		**Type III Sum of Squares**	**df**	**Mean Square**	**F**	**Sig.**	**Partial Eta Squared**
Distance	Level 1 vs. Level 6	423,801	1	423,801	54,522	,000	,773
	Level 2 vs. Level 6	347,437	1	347,437	82,192	,000	,837
	Level 3 vs. Level 6	195,886	1	195,886	72,390	,000	,819
	Level 4 vs. Level 6	94,180	1	94,180	63,067	,000	,798
	Level 5 vs. Level 6	18,498	1	18,498	11,864	,003	,426
Error(Distance)	Level 1 vs. Level 6	124,369	16	7,773			
	Level 2 vs. Level 6	67,635	16	4,227			
	Level 3 vs. Level 6	43,295	16	2,706			
	Level 4 vs. Level 6	23,893	16	1,493			
	Level 5 vs. Level 6	24,946	16	1,559			
Sentence	Level 1 vs. Level 5	76,424	1	76,424	31,747	,000	,665
	Level 2 vs. Level 5	51,888	1	51,888	12,468	,003	,438
	Level 3 vs. Level 5	99,903	1	99,903	14,858	,001	,481
	Level 4 vs. Level 5	30,624	1	30,624	4,839	,043	,232
Error(Sentence)	Level 1 vs. Level 5	38,516	16	2,407			
	Level 2 vs. Level 5	66,587	16	4,162			
	Level 3 vs. Level 5	107,582	16	6,724			
	Level 4 vs. Level 5	101,256	16	6,328			
Familiarity	Level 1 vs. Level 3	,514	1	,514	,105	,750	,007
	Level 2 vs. Level 3	10,177	1	10,177	5,571	,031	,258
Error(Familiarity)	Level 1 vs. Level 3	78,418	16	4,901			
	Level 2 vs. Level 3	29,227	16	1,827			

In table 4.5 the pairwise comparisons between the six distances are shown. According to the table, responses for near distances (namely .5m, 1m and 2m) are not significantly different from each other, while the distant ones (4m, 8m, and 16m) are significantly different. This overlap can be related to the logarithmic range of the distances and the overestimation of distance perception inside of the peripersonal space.

Table4.5: Output of Pairwise Comparisons of Distances: The numbers indicate distances respectively as; .5m, 1m, 2m, 4m, 8m and 16m

Pairwise Comparisons						
(I) Distance		**Mean Difference (I-J)**	**Std. Error**	**Sig.b**	**95% Confidence Interval for Difference b.**	
					Lower Bound	Upper Bound
1	2	-,472	,439	1,000	-1,984	1,040
	3	-1,598	,566	,182	-3,546	,349
	4	-2.639*	,724	,033	-5,132	-,146
	5	-3.950*	,521	,000	-5,745	-2,155
	6	-4.993*	,676	,000	-7,322	-2,664
2	1	,472	,439	1,000	-1,040	1,984
	3	-1.126*	,296	,023	-2,146	-,107
	4	-2.167*	,390	,001	-3,510	-,824
	5	-3.478*	,368	,000	-4,746	-2,209
	6	-4.521*	,499	,000	-6,238	-2,803
3	1	1,598	,566	,182	-,349	3,546
	2	1.126*	,296	,023	,107	2,146
	4	-1,041	,303	,051	-2,084	,002
	5	-2.351*	,281	,000	-3,318	-1,385
	6	-3.395*	,399	,000	-4,769	-2,020
4	1	2.639*	,724	,033	,146	5,132
	2	2.167*	,390	,001	,824	3,510
	3	1,041	,303	,051	-,002	2,084
	5	-1.311*	,377	,047	-2,610	-,011
	6	-2.354*	,296	,000	-3,375	-1,333
5	1	3.950*	,521	,000	2,155	5,745
	2	3.478*	,368	,000	2,209	4,746
	3	2.351*	,281	,000	1,385	3,318
	4	1.311*	,377	,047	,011	2,610
	6	-1.043*	,303	,050	-2,086	-4,538E-05
6	1	4.993*	,676	,000	2,664	7,322
	2	4.521*	,499	,000	2,803	6,238
	3	3.395*	,399	,000	2,020	4,769
	4	2.354*	,296	,000	1,333	3,375
	5	1.043*	,303	,050	4,538E-05	2,086
Based on estimated marginal means						
*. The mean difference is significant at the						
b. Adjustment for multiple comparisons: Bonferroni.						

Table 4.6 show the pairwise comparisons of the five sentences. The numbers indicate the sentences respectively as; "Dünya Gezegendir.", "Benden kazak aldı.", "Ormanda ağaç var.", "O, radyoyu açtı.", "Kaşık çekmecede." As it is shown in the table, the fifth sentence is significantly different from the others, except for sentence 4.

Table4.6: Output of Pairwise Comparisons of Sentences

Pairwise Comparisons

(I) Sentence		**Mean Difference (I-J)**	**Std. Error**	**Sig.b**	**95% Confidence Interval for Difference b.**	
					Lower Bound	Upper Bound
1	2	,373	,345	1,000	-,748	1,494
	3	-,304	,598	1,000	-2,249	1,641
	4	,778	,551	1,000	-1,014	2,571
	5	2.120*	,376	,000	,897	3,344
2	1	-,373	,345	1,000	-1,494	,748
	3	-,677	,423	1,000	-2,054	,700
	4	,405	,503	1,000	-1,230	2,039
	5	1.747*	,495	,028	,138	3,356
3	1	,304	,598	1,000	-1,641	2,249
	2	,677	,423	1,000	-,700	2,054
	4	1,082	,469	,347	-,443	2,607
	5	2.424*	,629	,014	,379	4,469
4	1	-,778	,551	1,000	-2,571	1,014
	2	-,405	,503	1,000	-2,039	1,230
	3	-1,082	,469	,347	-2,607	,443
	5	1,342	,610	,429	-,642	3,326
5	1	-2.120*	,376	,000	-3,344	-,897
	2	-1.747*	,495	,028	-3,356	-,138
	3	-2.424*	,629	,014	-4,469	-,379
	4	-1,342	,610	,429	-3,326	,642
Based on estimated marginal means						
*. The mean difference is significant at the						
b. Adjustment for multiple comparisons: Bonferroni.						

Table4.7: Output of Pairwise Comparisons of Familiarity

Pairwise Comparisons						
(I) Familiarity		Mean Difference (I-J)	Std. Error	Sig.a	95% Confidence Interval for Difference a.	
					Lower Bound	Upper Bound
1	2	-,948	,553	,317	-2,425	,530
	3	-,174	,537	1,000	-1,609	1,261
2	1	,948	,553	,317	-,530	2,425
	3	,774	,328	,094	-,102	1,650
3	1	,174	,537	1,000	-1,261	1,609
	2	-,774	,328	,094	-1,650	,102
Based on estimated marginal means						
a. Adjustment for multiple comparisons: Bonferroni.						

In table 4.7 the pairwise comparisons between three levels of familiarity are shown. F1; shows the responses of the listener to the voice of his/her couple while F2 and F3 representing the responses to the most similar voice and the least similar voice of his/her couple, respectively. As it is shown, there is no statistically significant difference between those levels of familiarity.

4.3 Discussion

In line with previous studies on auditory distance perception in humans, we found that participants' judgments of source distance for intensity-normalized speech sounds were generally poor. They judged the distance of familiar speech voice more accurately than other speech voices, replicating several prior assumptions that people are better at estimating the source distance of familiar speech [Bronkhorst and Houtgast, 1999], [Wisniewski et al., 2012]. This study extends past behavioral studies by showing that benefits of familiarity effect reflect the processing of familiar voices rather than lexical familiarity. In the experiments only the two factors known to be related to distance perception which are familiarity and direct-to-reverberant ratio. The former, we used recorded speech samples collected from different subjects and the latter, we controlled by measuring BRIRs at different distances. The distant-dependent variations in intensity level were preserved by calibrating the sound presentation level and likely contributed to participants' performance.

Overall, the statistical results indicate that the investigated level of familiarity, that of the interpersonal level, have a statistically significant effect on auditory distance perception. The decision to fit the data with power functions was based on past reviews of auditory distance perception [Zahorik, 2005] that used similar methods. Exponent and constant parameters from the fitted functions, which shows the amount of non-linear and linear compression or expansion of the functions, were, in most cases, similar to the past studies. The mean exponent from the Zahorik et al. (2005) and Flutt et al. [Flutti et al., 2013] review were similar (within one SD (standard deviation)) to that observed in auditory distance perception response condition.

The constant values for the data were somewhat higher than reported by Zahorik et al. (2005). This low R^2values may be caused by the variability between subjects in their usage of the response scale that lack a visual anchor and feedback; also they were completely naive in the aspect of attending auditory distance perception studies. Because dataset that the Zahorik et al. (2005) used was based on average values from different studies, issues that are related to individual subject variability were reduced, which may have also accounted for the somewhat higher average R^2values they reported shown in Table4.8.

Table4.8: Summary of results from past review of auditory distance perception study along with results from the current study.

Data Source	ADP Responses	F1	F2	F3	Zahorik et al. (2002)
α	0.4 ∓ 0.24	0.51 ∓ 0.04	0.34 ∓ 0.04	0.38 ∓ 0.04	0.39 ∓ 0.13
β	2.81 ∓ 0.1	2.2 ∓ 0.15	3.35 ∓ 0.22	3.01 ∓ 0.19	1.32 ∓ 0.56
R^2	0.3	0.26	0.14	0.18	0.6

Power function fit parameters and R^2 were presented to compare with the study of Zahorik et al. along with its Mean and S.E. values. F1, F2, and F3 represents the three level of familiarity

The overall response to the Sentence condition was unexpectedly or surprisingly effective on auditory distance perception. As it is demonstrated graphically in the Figure 4.4, the distance judgments for the fifth sentence (yellow line), shows that participants tended to respond to the distance by underestimating the sound source's real distance. Even if the participants listened to the fifth sentence from all other conditions randomly, they showed that they tended to perceive the sentence "Kaşık çekmecede" (The spoon is in the drawer) as if the source was nearby.

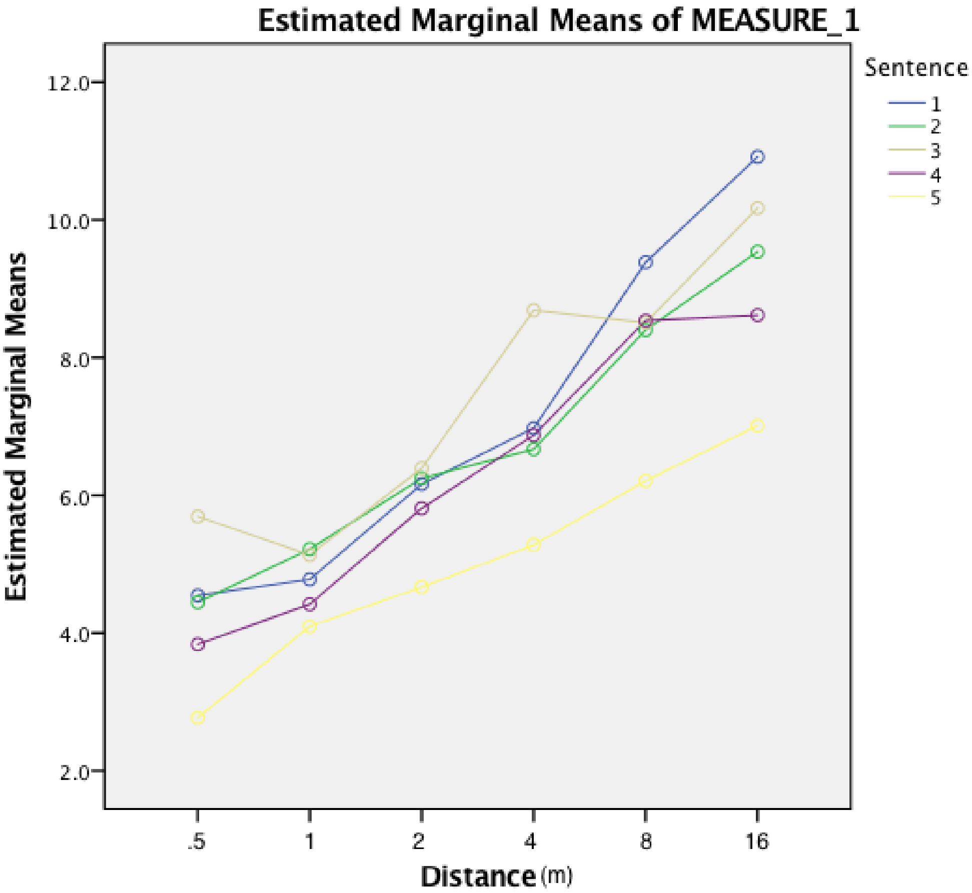

Figure 4.4: Power fits of Sentences

There can be many suggestions to explain this difference; for instance, the semantic features that represent a nearby location may be effective, or the voiced consonants (K, Ç, and Ş) may influence the emphasis of the sentence that causes the listener to perceive it closer. As it is assessed in Brungart's study, listeners require both acoustic and phonetic information to judge the distance from the production level of speech accurately. [Brungart and Scott, 2001]. If listeners use the phonetic information to distinct shouted speech from whispered speech, then it is possible to infer that the sentences used in this study phonetically evoked listeners to judge individual sentences as near or far sounds, whether their physical distances indicates the opposite.

Another question that comes to mind about the potential reason for why S5 is different than the others is; do low frequencies carry farther than high frequencies? The reason has to do with what's stopping the sound. If it weren't for attenuation (absorption) sound would follow an inverse square law as it is briefly mentioned in Chapter 2. Reflection is also frequency dependent. High frequencies are better reflected whereas low frequencies are able to pass through the barrier. The difference between S5 and others may cause from the reflection and attenuation differences between low and high frequencies of the syllables.

These findings should not be taken as evidence for the sentence condition has a huge effect on auditory distance perception but, rather it should be taken into consideration that speech as a cognitive factor has a contribution that should be further studied.

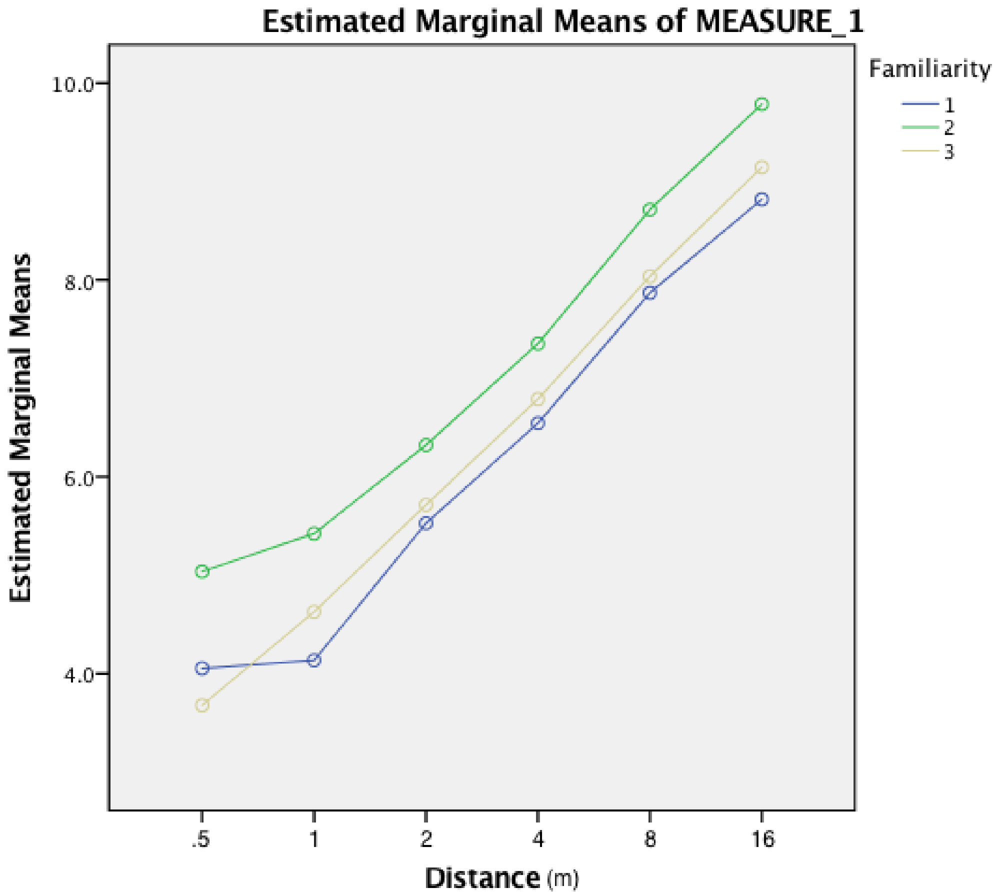

Figure 4.5: Power fits of Sentences

Johnsrude et al. state that although the fact of which aspects of voice segregation are affected by familiarity cannot be explained, the overall results show a substantial effect of familiar voice, regarding aiding speech recognition [Johnsrude et al., 2013]. We found that familiarity in a level of inter-personal voice has not a statistically significant effect, but as it can be seen on the graphic 4.5, for the most familiar voice people were show a tendency to give close distances, indicating that a familiar or reproducible sound (exposed to their real-life partner's voice every day) can have compressive effect on distance judgment. It may have shown a very strong cognitive effect that people are used to their partner's voice hearing from a near distance, such that they cannot give accurate judgments to F1 (partner's voice) when it's coming from further distances.

CHAPTER 5

CONCLUSION

Auditory distance perception is defined as a multidimensional and complex phenomenon that needs to be investigated extensively to enhance our understanding. In addition to the current knowledge and studies, auditory distance perception mostly remains unexplored compared with localization of sound source direction. There is still some lack of unexplained effects of the cognitive cues that provide the auditory distance judgment in humans. From this lack of knowledge, some questions arise, such as what are the effects of a familiar voice on auditory distance judgment, how it works and what are the limits of familiarity? To study those questions in a scientific way this thesis had the aim of exploring whether "Inter-personal voice familiarity would have a significant effect on auditory distance perception" as the Null Hypothesis.

A subjective experiment was conducted in order to understand whether a very specific familiarity condition can affect human distance perception. Participants were tested against their partner's voice vs. a similar voice and a dissimilar voice relative to their real-life partner's. Reverberation was the only acoustic cue besides of the familiarity to detect the distance. One of the themes that emerged from the analysis of interpersonal voice familiarity was interpersonal voice familiarity is not as effective as other kinds of familiarity aspects which defined in previous studies. Resulting from that possibly there is a lower limit considering familiarity as a cue. Normal speech considered as just a speech whether it contains a familiar voice or not. Surprisingly, the sentence condition had unexpected effects on auditory distance perception. It possibly indicating that the semantics or/and phonetics can serve as a cognitive cue. An interesting future study might involve focusing on two linguistic subgroups, namely the effects of semantic and phonetic aspects of the speech.

The overall conclusion of the previous studies on distance perception was that people are not able to accurately give distance judgment of a sound source. This outcome shows the need of considering this lack of accuracy when developing an auditory distance information; which means, perceived distance is typically compressed by comparing with physical distance. Also, recent studies suggest that auditory distance perception is enhanced by familiar sound sources in reverberant environments. Besides of speech and language familiarity, sounds that can be easily encountered in daily routine, that may be used in auditory interfaces to convey information. To do that, familiarity effect on auditory distance perception must be investigated more deliberately to understanding what causes a compression when we estimate the distance of the sound source.

As this study demonstrates, there is not sufficient understanding about the auditory distance perception and there are still gaps in our knowledge especially for the cognitive functions which play a role in auditory perception. In contrast to the lack of substantial knowledge about the area, this study offers suggestive evidence for that there may be a lower limit when investigating the familiarity effect on auditory distance perception. Hence, this leads to the following conclusion that more research is needed to conduct on the different levels of familiarity effect which may or may not affect the auditory distance judgments. In other words, a familiarity scale should be created to understand the upper and lower limits that what we understand from "familiarity".

This study is important because it represents a first attempt at identifying a clear familiarity scale that underlies judgments of auditory distance, and is thus limited in several respects. We directly measured participants' ability to localize sound sources with a continuous distance scale (0m-20m) and so we can assess whether their performance in this dichotomous auditory task accurately reflects their spatial acuity. This kind of free and continuous localization judgment caused a very low accuracy of distance perception, increased the standard deviation. Furthermore, it is possible that participants' used acoustic cues to differentiate sounds without perceiving them as spatial cues (i.e., they could distinguish the distance, but did not perceive the differences as corresponding to changes in the position of the source). Also, given that sound localization often occurs rapidly and involuntarily, participants' brains are likely continuously monitoring for the presence of such cues. Nevertheless, additional studies will be needed to definitively identify the auditory distance estimation techniques, as well as the factors that constrain the accuracy with which a particular individual can judge the distance to a sound source.

Findings from this study surprisingly suggest that phonetic or semantic information may be particularly relevant when intensity cues are not reliable indicators of source distance. This leaves the question of why phonetic processing might increase the availability of localization cues unanswered. One possibility is that familiar speech is processed more automatically, freeing brain resources for extracting auditory distance cues. Another possible factor, not considered in previous studies, is that some speech sounds are not only more familiar, they are also more reproducible (e.g. the fifth sentence). The fifth sentence represents a different affordance and proximity compared to the other sentences. If reproducible sounds or proximity meaning activate motor representations relevant to producing those sounds, then the availability of multimodal stimulus representations could enhance processing of acoustic cues.

As future work, this study revealed a substantial amount of research topics that need to be investigated to understand underlying effects of familiarity and linguistics on auditory distance perception. For instance, emotional aspects of the sentences or of the speaker, whether the sentence or the voice evokes any emotions and if it evokes emotions, whether are they effective on auditory distance perception or not, should be investigated as a future work. Secondly, selection of the semantic properties of stimuli must be controlled to avoid or apply some possible effects of semantic properties on the auditory distance perception.

Finally for the possible practical application for these findings it may be considered that human auditory system consists of complex contribution of different attributions when rendering a distance information through hearing; when creating a realistic virtual auditory environment, simulation of proper distance cues is important but as it can be deducted that not only the acoustic cues but also the cognitive ones are effective and that should be considered in order to create a realistic virtual auditory environment.

REFERENCES

[Anderson and Zahorik, 2014] Anderson, P. W. and Zahorik, P. (2014). Auditory/visual distance estimation: Accuracy and variability. *Frontiers in Psychology*, 5(SEP):1–11.

[Blauert and Butler, 1985] Blauert, J. and Butler, R. A. (1985). Spatial Hearing: The Psychophysics of Human Sound Localization. *The Journal of the Acoustical Society of America*, 77(1):334–335.

[Bronkhorst and Houtgast, 1999] Bronkhorst, A. W. and Houtgast, T. (1999). Auditory distance perception in rooms. *Nature*, 397(6719):517–20.

[Brungart and Rabinowitz, 1999] Brungart, D. S. and Rabinowitz, W. M. (1999). Auditory localization of nearby sources. Head-related transfer functions. *The Journal of the Acoustical Society of America*, 106(3):1465–1479.

[Brungart and Scott, 2001] Brungart, D. S. and Scott, K. R. (2001). The effects of production and presentation level on the auditory distance perception of speech. *The Journal of the Acoustical Society of America*, 110(1):425–440.

[Çağlayan and Özkurt, 2017] Çağlayan, E. and Özkurt, T. E. (2017). Classification of Emotions in Vocal Responses.

[Cooke, 1993] Cooke, M. (1993). *Modelling auditory processing and organisation.* Cambridge University Press, 6th edition.

[Edgell and Noon, 1984] Edgell, S. E. and Noon, S. M. (1984). Effect of violation of normality on the t test of the correlation coefficient. *Psychological Bulletin*, 95(3):576–583.

[Farina, 2007] Farina, A. (2007). Advancements in impulse response measurements by sine sweeps. pages 1–12.

[Flutti et al., 2013] Flutti, k. F., Mermagen, T., and Letowski, T. (2013). Auditory perception in open field: distance estimation. *Army Research Laboratory (ARL)*, ARL-TR-652(July).

[Gardner, 1969] Gardner, M. B. (1969). Distance estimation of 0° or apparent 0°-oriented speech signals in anechoic space. *The Journal of the Acoustical Society of America*, 45(1):47–53.

[Glasberg and Moore, 1990] Glasberg, B. R. and Moore, B. C. J. (1990). Derivation of auditory filter shapes from notched-noise data. *Hearing Research*, 47(1-2):103–138.

[Gogel and Tietz, 1973] Gogel, W. C. and Tietz, J. D. (1973). Absolute motion parallax and the specific distance tendency. *Perception & Psychophysics*, 13(2):284–292.

[Johnsrude et al., 2013] Johnsrude, I. S., Mackey, A., Hakyemez, H., Alexander, E., Trang, H. P., and Carlyon, R. P. (2013). Swinging at a Cocktail Party: Voice Familiarity Aids Speech Perception in the Presence of a Competing Voice. *Psychological Science*, 24(10):1995–2004.

[Kolarik et al., 2015] Kolarik, A. J., Moore, B. C. J., Zahorik, P., Cirstea, S., and Pardhan, S. (2015). Auditory distance perception in humans: a review of cues, development, neuronal bases, and effects of sensory loss. *Attention, Perception, & Psychophysics*, pages 1–23.

[Langendijk and Bronkhorst, 1999] Langendijk, E. H. and Bronkhorst, A. W. (1999). The contribution of spectral cues to human sound localization. *The Journal of the Acoustical Society of America*, 105(2):1036–1036.

[Lewicki, 2002] Lewicki, M. S. (2002). Efficient coding of natural sounds. *Nature Neuroscience*, 5(4):356–363.

[Loomis et al., 1998] Loomis, J. M., Klatzky, R., Philbeckandreginald, J. W., and Golledge, G. (1998). Assessing auditory distance perception using perceptually directed action. *Perception & Psychophysics*, 60(6):966–980.

[Mathôt et al., 2012] Mathôt, S., Schreij, D., and Theeuwes, J. (2012). OpenSesame: an open-source, graphical experiment builder for the social sciences. *Behavior research methods*, 44(2):314–24.

[Mershon and Bowers, 1979] Mershon, D. H. and Bowers, J. N. (1979). Absolute and relative cues for the auditory perception of egocentric distance. *Perception*, 8(3):311–322.

[Mershon and King, 1975] Mershon, D. H. and King, L. E. (1975). Intensity and reverberation as factors in the auditory perception of egocentric distance. *Perception & Psychophysics*, 18(6):409–415.

[Park, 2003] Park, A. (2003). Using the Gammachirp Filter for Auditory Analysis of Speech. pages 1–18.

[Patterson and Moore, 1986] Patterson, R. and Moore, B. (1986). *Auditory filters and excitation patterns as representations of frequency resolution.* Frequency Selectivity in Hearing. Academic Press Ltd, London.

[Patterson and Holdsworth, 1996] Patterson, R. D. and Holdsworth, J. (1996). A functional model of neural activity patterns and auditory images. *Advances in Speech, Hearing, and Language Processing*, 3:547–563.

[Pearson, 1931] Pearson, E. (1931). The test of significance for the correlation coefficient. *Journal of the American Statistical Association*, 26(174):128–134.

[Plenge, 1974] Plenge, G. (1974). On the differences between localization and lateralization. *The Journal of the Acoustical Society of America*, 56(3):944–951.

[Rungta et al., 2017] Rungta, A., Rewkowski, N., Klatzky, R., Lin, M., and Manocha, D. (2017). Effects of virtual acoustics on dynamic auditory distance perception. *The Journal of the Acoustical Society of America*, 141(4):EL427–EL432.

[Tervo, 2009] Tervo, S. (2009). Direction estimation based on sound intensity vectors. *European Signal Processing Conference*, (Eusipco):700–704.

[Valero and Alias, 2012] Valero, X. and Alias, F. (2012). Gammatone Wavelet features for sound classification in surveillance applications. *European Signal Processing Conference*, (Eusipco):1658–1662.

[Vesa, 2007] Vesa, S. (2007). Sound Source Distance Learning Based on Binaural Signal. *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pages 271–274.

[Wisniewski et al., 2012] Wisniewski, M. G., Mercado, E., Gramann, K., and Makeig, S. (2012). Familiarity with speech affects cortical processing of auditory distance cues and increases acuity. *PLoS ONE*, 7(7).

[Zahorik, 1996] Zahorik, P. (1996). Auditory distance perception: A literature review. *Phd Preliminary Examination, University of Wisconsin*, pages 1–13.

[Zahorik, 2002a] Zahorik, P. (2002a). Assessing auditory distance perception using virtual acoustics. *The Journal of the Acoustical Society of America*, 111(4):1832–1846.

[Zahorik, 2002b] Zahorik, P. (2002b). Direct-to-reverberant energy ratio sensitivity. *The Journal of the Acoustical Society of America*, 112(5 Pt 1):2110–2117.

[Zahorik, 2005] Zahorik, P. (2005). Auditory Distance Perception in Humans: A Summary of Past and Present Research. *Acta acustica united with Acoustica*, 91(3):409–420.

[Zwislocki, 2009] Zwislocki, J. J. (2009). *Sensory Neuroscience: Four Laws of Psychophysics.* Springer Science+Business Media, LLC 2009, Syracuse, New York.

Appendix A

MATERIAL DETAILS

A.1 Hallway

Figure A.1: Dummy head microphone Neumann KU 100

Figure A.2: The hallway used for recording binaural room impulse responses

Appendix B

MATLAB CODE

B.1 Comparing Recorded Sounds Using Fourth Order Gammatone Filter

We propose to use Gammatone filtering functions, which are well known for their application to human auditory modeling, especially to model the cochlear frequency response. They are asymmetric and have a variable duration that depends on their central frequency. Thus, filtering a signal with a Gammatone filterbank is similar to a Wavelet transform in the sense that all basis functions are scaled versions of the mother function at the first central frequency [Valero and Alias, 2012] .

The Gammatone filter takes its name from the impulse response g(t,B) (see Fig. 1), which is the product of a Gamma distribution function and a sinusoidal tone centered at the f_c frequency, being computed as;

$$g(t, B) = K\ t^{(n-1)} e^{-2\pi Bt} cos(2\pi\ f_c t + \varphi) \quad t > 0$$

where K is the amplitude factor; n is the filter order; f_c is the central frequency in Hertz; φ is the phase shift; and B represents the duration of the impulse response. The scaling of the proposed Gammatone Wavelet function is controlled by B, which is related to the Equivalent Rectangular Bandwidth (ERB), a psychoacoustic measure of the auditory filter width at each point along the cochlea [Patterson and Holdsworth, 1996].

Patterson et al. (1992) show that the impulse response of the gammatone function of order 4 provides an excellent fit to the human auditory filter shapes derived by Patterson and Moore (1986) [Patterson and Moore, 1986]. Glasberg and Moore (1990) have summarized human data on the equivalent rectangular bandwidth (ERB) of the auditory filter with the function [Glasberg and Moore, 1990]:

$$ERB = 24.7(4.37\ 10^{-3} f + 1)$$

When the order of the filter is 4, the bandwidth B of the gammatone filter is 1.019 ERB. The filter-bank is normally defined in such a way that the filter center frequencies are distributed across frequency in proportion to their bandwidth, known as the ERB scale [Glasberg and Moore, 1990]. The ERB scale is approximately logarithmic, on which the filter center frequencies are equally spaced [Park, 2003].

This gammatone filter implementation is based on Martin Cooke's Ph.D. study (Cooke, 1993) using the baseband impulse invariant transformation [Cooke, 1993]. In this study, audio classification provided with the filtered sounds by comparing their spectral powers. The Matlab code generates and lists of two audio files in each loop, those files correspond to the most and the least similar audio files to the source audio. By this classification, the three levels of familiarity determined; F1 (source audio), F2 (the most similar audio to the source), F3 (the least similar audio to the source).

Free source code: gammatone.cc <http://www.gnu.org/licenses/> The following code used for computing gammatone filter function and correlating sounds to generate familiarity levels;

```
clear all;
close all;
clc;

fs = 44100;
sentenceCount = 5;
coupleCount = 12;
genders =  female ,  male  ;
prefix =  Couple ;
extension =  .wav ;

 constructing sound names
for sentence=1:sentenceCount

for genderIndex = 1:2
    gender = genders(genderIndex);

    soundNames =  ;
    for couple=1:coupleCount
        soundName = strcat(prefix, num2str(couple), char(gender), num2str(sentence), extension);
        soundNames = [soundNames; soundName];
    end

maxIndex = length(soundNames);         Max Index for sound names
fs = 44100;                            Sampling rate of files
lowcf = 50;                          lowest frequency
highcf = fs/2;                      high frequency
numchans = 64;

 To construct the gammatonefilter, erb coefficients are used as the most accurate coefficients for an auditory data.
 Therefore, erb coefficients are calculated
 "When the order of the filter is 4, the bandwidth b of the gammatone filter is 1.019 ERB."
 REFF: http://staffwww.dcs.shef.ac.uk/people/N.Ma/resources/gammatone/

cfs = makeErbCFs(lowcf,highcf,numchans);
```

```

disp( Compare Sounds Using Fourth Order GammaTone Filter );

  Fetching sounds

samplingFrequencies = zeros(maxIndex,1);
maxLengthOfAudios = 0;    Max audio length will be hold to compare sounds

  Finding Max Sound Length

disp( Starting to read files and finding max file length );
for i = 1: maxIndex
    soundName = char(soundNames(i));
    y = audioread(soundName);
    lengthY = length(y);
    if maxLengthOfAudios   lengthY
        maxLengthOfAudios = lengthY;
    end
end
disp( Fetched max file length );

  Constructing zero arrays for sound data

originalSoundDatas = zeros(maxLengthOfAudios, maxIndex);
filteredSoundDatas = zeros(maxLengthOfAudios, maxIndex);
longtspec = zeros(maxIndex, 64);

  In this for loop gammatonefilters are created using erb coefficients for the filter.

disp( Reading files );
for i = 1: maxIndex
    soundName = char(soundNames(i));
    disp([ Reading File:    soundName]);
    y = audioread(soundName);
    y = filterA(y, fs);
    lengthY = length(y);
    originalSoundDatas(1:lengthY,i) = y;
    disp([ Creating gammatone filter for File:    soundName]);
    y filtered = gammatoneFast(y, cfs,fs,true);                              Sound filtered with gammatoneFast algorithm
    lengthY filtered = length(y filtered);
    ysspect = sum(y filtered . 2,1);    HH
    filteredSoundDatas(1:lengthY filtered,i) = y filtered(:,numchans);    Filtered sounds are stored in multidimensional array
    longtspec(i,:) = ysspect;
```

```
end
disp( Finished Fetching sounds and Creating Filtered Sounds );
disp( Measuring correlation between sounds );

for ind = 1:maxIndex
    longtspec(ind,:)=longtspec(ind,:)./norm(longtspec(ind,:));
end

for ind = 1:maxIndex
    for jnd = 1:maxIndex
        sMat(ind,jnd) = dot(longtspec(ind,:), longtspec(jnd,:));
    end
end

max = 0;
min = 1;
tolerance = 0.0001;
for ind = 1:maxIndex
    for jnd = 1:maxIndex
        val = sMat(ind,jnd);
        if abs(val   1)   tolerance    val    max
            max = val;
            maxInd = [ind jnd];
        end

        if val   tolerance    val    min
            min = val ;
            minInd = [ind jnd];
        end
    end
end

genderStr = char(gender);

disp([ Index for sound   genderStr   sentence:   num2str(sentence)]);
disp( Max Index )
maxInd
disp( Min index )
minInd


figure;

imagesc(sMat);
title([genderStr   sentece:   num2str(sentence)]);

end
```

```
end
```

B.2 Convolving Sounds with BRIRs

```
clear all;
close all;
clc;

disp( Starting dist conv );
baseDirLocation = /Users/ozgendk/Documents/MATLAB ;
soundDir = [baseDirLocation /morphed ];
outputDirLocation = [baseDirLocation /output ];
HRIRDirLocation = [baseDirLocation /HRIRs ];

HRIRDir = dir(HRIRDirLocation);
HRIRDir = leandirname(HRIRDir);

participants = dir(soundDir);
participants = leandirname(participants);
for participantIndex = 1:length(participants)
     participantIndex = 1;  Only for test
    participantCell = participants(participantIndex);
    participant = participantCell 1 ;
    disp([ Processing Participant  participant]);
    participantDir = [soundDir /  participant];
    morphedVoicesForParticipant = dir(participantDir);
    morphedVoicesForParticipant = leandirname(
        morphedVoicesForParticipant);
    for morphedVoiceIndex = 1: length(
        morphedVoicesForParticipant)
          morphedVoiceIndex = 1;  Only for test
        morphedVoiceCell = morphedVoicesForParticipant(
            morphedVoiceIndex);
        morphedVoiceStr = morphedVoiceCell 1 ;

        for HRIRIndex = 1: length(HRIRDir)
             HRIRIndex = 1;  Only for test
            HRIRCell = HRIRDir(HRIRIndex);
            HRIRStr = HRIRCell 1 ;
            dirLocation = [outputDirLocation /  participant];
            if(exist(dirLocation,  dir ) == 0)
                mkdir(dirLocation);
            end
            convolvewithhrir(HRIRStr, HRIRDirLocation,
                morphedVoiceStr, participantDir,
                outputDirLocation, participant);

        end
```

```
    end
end
```

B.3 Experimental GUI / Slider Screen

```
mycanvas = canvas()
mymouse = mouse(timeout=10)
mykeyboard = keyboard(keylist=[ space ], timeout=50)

    Set slider dimensions. This assumes that 0,0 is the
      display center, which is
    the default in OpenSesame = 3.
slider w = 10
slider h = 500
slider x =  slider w/2
slider y =  slider h/2
MAXDISTANCE = 20

   ynorm  =1
   while True:
 Determine the slider fill based on the mouse position
        pos, time = mymouse.get pos()
        x, y = pos
        slider fill = y
        if(y     slider y):
            slider fill =  slider y
        if(y    slider y):
            slider fill = slider y
        slider fill += slider y
  slider fill = max(slider fill ,   slider y)
  slider fill = max(slider h , min(0, slider y) y)
        mycanvas.clear()
          Draw some text (this can be anything)
        center = 305
        r   = 30
        top = center    r
        mycanvas.circle(0, center , r, fill=False, color=
           white )
        mycanvas.line(10,top,0,top 10)
        mycanvas.line( 10,top,0,top 10)

        mycanvas.text("Dinledigin ses ne kadar uzakliktan
           geldi?", y=slider y 100)
        mycanvas.text("Secmek icin tiklayin", y=slider y 70)
 Draw the slider frame
        mycanvas.rect(slider x , slider y , slider w , slider h)
 Draw the slider fill
        mycanvas.rect(slider x ,  slider y , slider w  ,
           slider fill , fill=True)
```

```
  Draw the mouse cursor
  mycanvas.arrow(x+5, y+10, x, y)
        ynorm =  (( sliderfill    MAXDISTANCE 1.0)/sliderh)
        ynorm  = float(format(ynorm,   .1f ))
        mycanvas.text(text=str(ynorm)+m , x=x, y=y)
        mycanvas.show()

  Poll the mouse for buttonclicks
        button, position, timestamp = mymouse.getclick()
        if button is not None:
            mycanvas.text("Uzaklik secildi     SPACE tusuna
               basip devam ediniz", x=85, y=slidery+200)
            mycanvas.show()
            break
    while True:
        key, timestamp = mykeyboard.getkey()
        if key is not None:
            break
   Set the slider response as an experimental variable
```

Appendix C

POWER LAW FITTING

C.1 Constant and power-law exponent codes and values for all Participants

```
sonuc=data.frame()
sub=1
while(sub 22)

birinciT=0
ikinciT1=0
ikinciT2=0
ustkat=0
altkat=0
altkat1=0
altkat2=0
beta=0

alpha1=0
alpha2=0
alphason=0

for(i in 1:90)
x=ozg[ozgsub==sub,][i,"dist"]
y=ozg[ozgsub==sub,][i,"Mres"]
   birinciT=birinciT+log(x) log(y)
   ikinciT1=ikinciT1+log(x)
   ikinciT2=ikinciT2+log(y)

   altkat1=altkat1+log(x) log(x)
   altkat2=altkat2+log(x)

   alpha1=alpha1+log(y)

ustkat= 90 birinciT ikinciT1 ikinciT2
altkat =90 altkat1 altkat2 altkat2
beta=ustkat/altkat
alphason=(alpha1 beta altkat2)/90

```

```
print(c(beta,alphason))

sonuc=rbind(sonuc,c(beta,alphason))
sub=sub+1

names(sonuc)=c("beta","alpha")
```

The following figure shows the results of the power fitting calculation, where the beta values indicate the non-linear compression and the alpha show the constant value.

	beta	alpha
1	0.51401549	0.73377614
2	0.48676886	0.07931369
3	0.14129878	1.33763309
4	0.50772900	1.34588610
5	0.28063324	0.98866182
6	0.11726573	1.58649673
7	0.24914292	2.24940235
8	0.31082513	1.79077591
9	0.23932338	0.68926206
10	0.42971849	0.85583566
11	0.49587759	0.78352411
12	0.11147313	1.13164390
13	0.23342486	1.83820158
14	0.33918733	0.84920291
15	0.40290626	1.05269257
16	0.02898793	1.15100937
17	0.01118143	-0.74697705
18	0.09993094	1.77352322
19	0.08989698	1.75249799
20	0.16460257	1.29159141
21	0.38970053	0.28306975

Figure C.1: All 21 participants' non-linear compression values

The following code plots the graphics of the power fits for all the participants.

```
  for(i in 1:21) ggsave(filename=paste(as.character(i),".jpg",
     sep = ""),
                              plot=ggplot(data=ozg[ozg sub==i,],aes(x=
                                 dist,y=Mres)) +
                                geom point(size=0.2)+stat function(fun
                                   =function(x) exp(sonuc alpha[i]) (
                                  x sonuc beta[i]) )+
                                stat summary(fun.y = function(a) prod(
                                  a) (1/length(a)), geom = "point",
                                  shape=0,size=3)+ scale x continuous
                                  (breaks = c(1,10))+
                                  scale y continuous(breaks = c(1,10)
                                  )+
                                ylab("Estimated Distance(m)")+xlab("
                                  Source Distance(m)")+coord trans(x
                                  ="log10",y="log10"),width = 4,
                                  height = 4)

birinciT=0
ikinciT1=0
ikinciT2=0
ustkat=0
```

```
altkat=0
altkat1=0
altkat2=0
beta=0

alpha1=0
alpha2=0
alphason=0


for(i in 1:length(sendat5Mres))
  x=sendat5[i,"dist"]
  y=sendat5[i,"Mres"]
  birinciT=birinciT+log(x) log(y)
  ikinciT1=ikinciT1+log(x)
  ikinciT2=ikinciT2+log(y)

  altkat1=altkat1+log(x) log(x)
  altkat2=altkat2+log(x)

  alpha1=alpha1+log(y)

ustkat= length(sendat5Mres) birinciT ikinciT1 ikinciT2
altkat=length(sendat5Mres) altkat1 altkat2 altkat2
beta=ustkat/altkat
alphason=(alpha1 beta altkat2)/length(sendat5Mres)

print(c(beta,alphason))

sonuc=rbind(sonuc,c(beta,alphason))


names(sonuc)=c("beta","alpha")
```

C.2 Constant and power-law exponent values for Familiarity and Sentences

```
22 0.33984295  0.96128116
23 0.27152085  1.29160294
24 0.34402912  1.08011057
25 0.39071165  1.10953047
26 0.28592601  1.21272843
27 0.26418130  1.39128016
28 0.37327125  1.05792092
29 0.27823135  0.78353113
> |
```

Figure C.2: The first three values show the familiarity based power fitting values while the last five values showing the sentences'