UNIVERSITY OF ALBERTA

Listener Discrimination Between Age Differences in Two Speakers: A comparison task involving recordings of the voices of Terry Gross and Ira Glass from different years

BY

Lara Ozdogan

A THESIS

SUBMITTED TO THE FACULTY OF ARTS

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF BACHELOR OF ARTS

DEPARTMENT OF LINGUISTICS

EDMONTON, ALBERTA

April, 2023

UNIVERSITY OF ALBERTA

FACULTY OF ARTS

UNIVERSITY OF ALBERTA

FACULTY OF ARTS

The undersigned certify that they have read and recommend to the Faculty of Arts for acceptance, a thesis entitled Listener Discrimination Between Age Differences in Two Speakers: A comparison task involving recordings of the voices of Terry Gross and Ira Glass from different years, submitted by Lara Ozdogan in partial fulfillment of the requirements for the degree of Bachelor of Arts.

Dr. Benjamin V. Tucker...................................... Honors Thesis Supervisor

Dr. Jorge Emilio Rosés Labrada.................................... For the Department

ABSTRACT

Often when listeners hear a speaker, they have an impression of the age of the speaker. The present research seeks to determine how well listeners can identify a speaker's age when provided with two different recordings from one speaker. Previous research has asked participants to place a speaker's age on a continuum or estimate the age of the speaker. In the present study, I use an AX task which asks listeners to determine whether the X recording is younger or older, from speech recorded from two speakers spanning approximately 20 years. I also asked a few questions about the listeners' language background, age, and experience with populations over the age of 40. I analyze the discrimination responses of our listeners and find that listeners are not much better than chance in determining age differences in the recordings. I do find that as the number of years between recordings increases, participants are more accurate at determining the age difference between recordings. Listeners have specific ideas about what acoustic features they are using to perform their discrimination; however, these do not necessarily lead to more accurate discrimination.

ACKNOWLEDGMENTS

I would like to express my gratitude to my honors supervisor, Dr. Benjamin V. Tucker for his support, guidance, and patience, and for giving me the opportunity to be a part of the Alberta Phonetics Laboratory and to start my research journey. I would also like to the University of Alberta Undergraduate Research Council for funding the collection of the stimuli used in this project through the Roger S. Smith Undergraduate Student Researcher Award. Additionally, I'd like to thank Tyler Schnoor for setting up my experiment for me, and to the rest of the members of the Alberta Phonetics Laboratory for their support, encouragement, and feedback throughout this process.

TABLE OF CONTENTS

TABLES AND FIGURES

CHAPTER 1. INTRODUCTION & BACKGROUND

As we age, many parts of our physical characteristics change. There are the characteristics we know are shifting, like our hair, or our sight, or our ease of movement. In addition, there is one aspect of our physical beings that is shifting, potentially without our awareness even, that characteristic being our speech. The present study seeks to expand research on the topic of aging speech, by investigating the interplay between shifting acoustic characteristics of speech along with age, and a listener's perceived age of a speaker. In this thesis, 1 study and 2 experiments will be described. The study involves an acoustic analysis of two speakers' voices, these speakers being Terry Gross and Ira Glass. The 2 experiments involve AX comparison tasks of the voices of Terry Gross and Ira Glass respectively. The ensuing sections will go through an exploration of the relevant literature (Chapter 1, section 1.2), along with my research questions and hypotheses, as well as a full description of the study and the 2 experiments separately, starting with Chapter 2: Acoustic Analysis, followed by Chapter 3: Experiment 1: Terry Gross, and Chapter 4: Experiment 2: Ira Glass. Each of the chapters involving the study and 2 experiments involves a breakdown of their respective methodologies, results, and discussion.

The first experiment involves an acoustic analysis of two speakers over the course of the lifespan. Although previous work has demonstrated how in general; the human voice shifts with age, much of the previous work has not looked at how the voice of **one** individual shifts over their lifespan. The ensuing two experiments are discrimination tasks in which listeners are asked to discriminate age differences between two recordings from the same speaker from two points in the speaker's lifespan. In these experiments, participants listen to two short recordings from one speaker, from two different years of the speaker's lifespan, and are then asked to determine which recording has the older voice. This task remains the same in both Experiment 1 and Experiment 2, the only difference is that Experiment 1 uses the voice of Terry Gross, while Experiment 2 uses the voice of Ira Glass.

1.1 How does speech change with age?

1.1.1 Anatomical and acoustical changes

Previous literature with regard to speech changes with age has found that certain acoustic characteristics of male voices, such as fundamental frequency and speaking rate increase and decrease with age respectively (Harnsberger et al. 2008). Other studies have determined that both amplitude and fundamental frequency decrease over the course of an individual's lifespan no matter the gender of the speaker (Reubold, Harrington, and Kleber 2010). These acoustic changes can be credited to anatomical and physiological changes in various parts of the mechanisms of speech. For example, according to Linville (1995, 183), as the body ages, the thorax stiffens and there is a decreased force of the respiratory muscles, which contributes to changes in airflow through the system. Other anatomical changes mentioned by Linville (1995, 183) include changes in the larynx such as "ossification and calcification of cartilage, and atrophy of muscle tissue." Essentially, there is an overall weakening of the phonatory system as well as the respiratory system. It is this weakening in the overall mechanisms that produce speech that result in acoustic changes. In the next subsection I will be discussing the prior work done on acoustic shifts over the lifespan, as well as the perceptual salience of these changes .

The acoustical features involved in speech analysis include features such as fundamental frequency (pitch) and formant frequencies. Although there are other features that are analyzed such as speech rate, amplitude, duration and pauses, for the purposes of the acoustic analysis performed in this thesis, fundamental frequency and formant frequency were analyzed. Previous literature regarding changes of fundamental frequency and formant frequencies with age has shown several shifts. For instance, Brown (1991) found that fundamental frequency has a “U” shaped curve. It increases, dips, then increases again in the later years of life. This finding has been supported in research of other individuals (Hunter et al.(2012), Reubold et al. (2010) and Stathopoulos (2011)). Many of these studies compare the voices of different speakers to determine how acoustic characteristics shift with age. The acoustic analysis study performed in the present work is used to confirm the above findings are consistent with the changes that occur through the lifespan of a single speaker.

1.2 Prior work

Much of the work related to aging speech has to do with determining (1) how our voice changes with age and (2) if listeners can estimate the age of speakers (typically by organizing them on a continuum), and (3) if listener characteristics influence accuracy of age discrimination. The following 3 subsections will look at prior work in these areas.

1.2.1 Production

Previous work performed on age-related shifting of acoustic characteristics has looked at various things; however, for the purposes of this thesis I will only discuss work on fundamental frequency, and the first and second formant frequencies.

A study by Fisher and Linville (1985), in which 75 women aged 25 to 80 were asked to make various vowel sounds showed that as age increases fundamental frequency (F0), F1, and F2 decrease. Other studies like that of Reubold, Harrington, and Kleber (2010) comfirmed Fisher and Linville’s (1985) finding, and expanded on it. In particular, upon looking at a 50 year span for two speakers, Reubold, Harrington, and Kleber (2010) found that as age increased, there was a decrease in both F0 and F1 for a female speaker (consistent with Fischer and Linville (1985)); however, they also found a V-shaped pattern in both F0 and F1 for the male speaker they analyzed. This V-shaped pattern means a decrease in F0/F1 followed by an increase at some point in the lifespan. Their results showed that for the male voice, the increase started in the late years, at around age 85 (Reubold, Harrington, and Kleber 2010, 642).

There is a common thread in the work that F0 decreases as age increases; however, there have been other studies, like that of Stathopoulos, Huber, and Sussman (2011), which have shown that fundamental frequency has a high variability in both young and old speakers (based on voice recordings from 192 male and female participants 4-93 years of age). Stathopoulos, Huber, and Sussman (2011) also found the aforementioned V-shape in F0 in both male and female participants, where in males F0 declined until about 50 years of age and then began to rise, and in females F0 decreased until about 60 years of age and then started to rise (which is different from the results found by Reubold, Harrington, and Kleber (2010)). For a more in depth review of the literature reviewing acoustic changes in speech over the lifespan, see Tucker, Ford and Hedges (2021).

1.2.2 Perception

As mentioned previously, much of the prior published work has looked at how individuals perceive acoustic changes in our voice. Said work has relied on a few types of experiments. One type involves a task where individuals are asked to listen to two recordings and choose which is the older voice. Others involved asking individuals to estimate the age of a speaker or to place speakers into age categories (ex. young, middle, old), or on an age continuum. In the following paragraph, I discuss each of these types.

An experiment performed by Bruckl (2007), which involved female voices recorded once, then 5 years later. The goals of this study involved determining how the female voice changes, and which characteristics contribute to the sound of “older” voice (Bruckl 2007, 3). According to the study, in 5 years female voices (aged 26, 27, 27, 30, 39, 40, 61, 67, and 87) shifted enough that participants doing an AxB (which involves a listener hearing recording “A” and then recording “B” and comparing them based on a provided question) perception task were able to determine which sample was from the speaker when they were older “2⁄3” of the time (Bruckl 2007, 3). Bruckl (2007) states that listeners responded to duration, tempo, pitch, hoarseness/ roughness/ breathiness, vocal tremor, and vowel quality in the speech sample to make their discriminations.

In another experiment, Ferguson et al. (2013) asked a group of listeners (17-50 years old) to categorize a voice’s age into one of three age categories (50-66, 67-83, 84-100) or estimate the actual age. The voices spanned from age 48-98 (48 years). According to the results, listeners were more accurate if the speaker was older than 68; however, the speaker age was consistently underestimated by around 5 years (Ferguson et al. 2013).

In an experiment where listeners placed speakers along an age continuum, Ryan and Burk (1974) asked female listeners to listen to male voices (aged between 40-80) and place them along an age continuum. Ryan and Burk (1974) found that listeners were able to accurately place speakers along an age continuum.

Overall, they found that in order to distinguish “older” speech, listeners use “air loss, voice tremor, imprecise consonants, and slow rate of articulation” (Ryan and Burk 1974, 187) to estimate age. In subsequent research performed by Ryan and Capadano (1978) mentioned that listeners said they responded to “pitch, tone, and volume of the voice as well as the clarity and speed” of the voice in their discriminations (Ryan and Capadano 1978, 3).

For a more in depth review of the literature looking at the perception of age in voice, see Hunter, Ferguson and Newman (2016) as well as Tucker, Ford and Hedges (2021).

1.2.3 Effect of listener characteristics on accuracy

Many other studies have focused on listener characteristics as variables for accuracy in discrimination. For example, Huntley, Hollien, and Shipp (1987) aimed to determine if listener age has an impact on the accuracy of their estimation of a speaker’s age. The results of their experiment showed that listeners were overall accurate in estimating speaker age; however, “ages of the younger talkers were overestimated, and the ages of the older talkers were underestimated” (Huntley, Hollien, and Shipp 1987, 3). In general, the listener groups performed similarly overall; however, the 20- and 30-year-old groups were different from others. These authors posited that younger listeners underestimated ages because they are just simply way younger than everyone, and the older speakers (over 70) overestimated younger voices since everyone is younger than them. In other words, fine discriminations can be more difficult if you are substantially younger or older than the other group (Huntley, Hollien, and Shipp 1987, 3).

A study performed by Moran, McCloskey, and Cady (1995) set out to determine if a listener’s race has any effect on the accuracy of age estimation of a speaker. They also sought to determine if there were differences between races in age related speech changes. The results of the study demonstrated that both “African-American and Caucasian listeners underestimate the age of Caucasian speakers” (Moran, McCloskey, and Cady 1995, 3). Their results further demonstrated that there were some differences between groups within certain acoustic characteristics; for example, the reading rate of speakers decreased with age for both racial groups (Moran, McCloskey, and Cady 1995, 3).

1.3 Research questions

Based on the literature, my goal is to determine the accuracy of a listener using an AxB experiment with both a male and female speaker over the course of 20 years of their lifespans. Previous research has not had listeners listen to large spans of speech from a male and female's life. My other aim is to determine how characteristics of a listener impact accuracy during this task.

Another aim of this research is to explore how the acoustic characteristics such as fundamental frequency and formant frequency shift over the lifespan of two speakers (one male, one female), and to determine if individuals listening to two recordings of the same speaker from two different years can determine which recording is from the speaker when they are older. Other aims of this project include determining if listener factors such as age, sex, native language, and frequency and enjoyment of time spent with individuals over 40 impact their accuracy on the listening task. Since research performed by Bruckl (2007) shows that listeners are able to distinguish age between recordings of a speaker 5 years apart, I hypothesize that listeners will perform well on this task, as long as the recordings have a difference of at least 5 years.

CHAPTER 2. ACOUSTIC ANALYSIS

This part of the thesis involves a corpus analysis, specifically an acoustic analysis of speech samples collected during the Summer of 2022 from May until August funded by a Roger S. Smith Undergraduate Research Award. The goal of this section of the thesis is to determine which acoustic features of a male and female speaker change over their respective lifespans. There were no participants for this experiment, as it involved internet data collection, including finding speakers with recorded, downloadable speech samples that stretch over a long period of time. The process to find the data was difficult; however, and the following sections will explore the process further.

2.1 Method

This corpus analysis required North American English data that spanned over the course of at least two decades, was publicly available, downloadable and had accessible transcripts. The goal was to find speakers with a span of speech covering at least two decades.

Initially, the speech of individuals such as Don Cherry and Christopher Plummer was going to be used, both of whom have long histories of television appearances. However, the speech of these individuals is not consistently recorded (year-to-year), nor are they downloadable with access to transcripts. The National Public Radio was the database which ended up providing the data for this analysis, as it has archived episodes of most of their produced shows. These archives also provide downloadable MP3s of each show, as well as transcripts. The speakers used in this analysis were Ira Glass (on NPR, “This American Life” from 1995-2022) and Terry Gross (on NPR, “Fresh Air” from 2000-2022).

2.1.1 Data collection & editing

For Terry Gross, I took two recordings per year, one from January and one from July of each year (each recording was approximately 5 minutes in length). Her recordings and transcripts were all downloaded from the NPR archives (which can be found here: https://www.npr.org/series/483233554/npr-selected-archives-two). For the purposes of this project, 22 of the recordings, including January and July of 2000 up to January and July of 2022 were used.

Upon completing data collection of Terry Gross, I collected the stimulus for Ira Glass, who has 26 years of recorded data. He has been the host of This American Life since 1996 (he was 36 when he started and is now 63), and all his recordings are available for download on the National Public Radio archives as well as on the archives of This American Life (the archives can be found at: https://www.thisamericanlife.org/archive). For Ira Glass I chose to collect 6 recordings from each year of the show (unlike Terry Gross see below), while aiming for 5-minute chunks of speech in each recording (as with Terry Gross). However, because of the nature of This American Life, the average amount of time per recording is around 2-3 minute. The challenge with his show is that he plays music throughout much of the show, and most of the speaking is produced by the individual he is interviewing, leaving little uninterrupted Ira Glass speech. It is for this reason I chose to use more recordings per year for Ira Glass than for Terry Gross. The way that I chose which recordings to use from each year was by doing every second month of the year (specifically, Jan, Mar, May, Jul, Sept, Nov.). On occasion, there was a rerun (for example, “Fiasco!” aired both in November 2013 and April 1997), or Ira Glass was sick, and consecutive months would have to be used. I also avoided using any clips with any background noise.

2.1.2 Materials

The materials used in this experiment were the recordings taken from Terry Gross (22 recordings) and Ira Glass (25 recordings), as well as the transcripts coinciding with each recording. Although originally 6 recordings per year were taken for Ira Glass, only one recording per year was used in the analysis, to be even with the data available for Terry Gross (who had only 1 recording per year available). The recordings used in the analysis were taken from January (for Ira Glass) and July (for Terry Gross) of each year. After downloading the full recording, I cut out 5-second-long intervals from each recording, ensuring the segment did not have long pauses, or noises like coughs.

2.1.3 Procedure

After collecting the data, the next step was to convert the MP3s downloaded from Terry Gross and Ira Glass into WAV files using a Praat script, as this is the format which works more effectively for data extraction. A possible caveat with this study is using MP3. Once you have compressed a recording into WAV, you cannot regain the information you may have lost in compressing the file, even if you return the file to an uncompressed format. A preliminary forced alignment using P2FA (Yuan and Liberman 2008) was run to determine which words in the transcripts were not already in the alignment dictionary.

With the list of missing words, I edited the transcripts and created new dictionary entries. For example, some “words” were in fact just dashes or spaces that needed to be deleted, or words spelled wrong that needed to be edited. The new dictionary entries were transcribed using the Arpabet (Klautau 2001). After the transcripts were edited for both Ira Glass and Terry Gross, and words that were missing from the dictionary were transcribed and added, the forced alignment could be run to create force aligned text grids with the spectrograms (see Appendix A, Example A.1). The text grids from this output are used in the extraction of acoustic measures.

The next step of the procedure was to extract the acoustic measures that will be used in the analysis of the collected data. A custom Praat script was used (for both speakers) to extract fundamental frequency (F0), F1, F2, F3, along with word and vowel duration. While F3 was extracted, we chose not to analyze it in the present investigation.

After these values were extracted, I used the R statistical programming language (R Core Team 2022) to investigate trends in the data and determine whether and how the selected acoustic measures shift over the life course of both Terry Gross and Ira Glass.

2.2 Results

To determine the relationship between acoustic characteristics, I completed a linear regression for each acoustic measure as a function of year per speaker.

2.2.1 Results for the acoustic analysis of Terry Gross

The first acoustic characteristic that I modeled as a function of year was the fundamental frequency of Terry Gross's voice. Figure 1a is a plot showing the modeled effect of year on the fundamental frequency of Terry Gross over time for the investigated time span. The actual effect size of year on Terry Gross' fundamental frequency is a decrease of 10.51 Hz. Figure 1b is a linear regression which demonstrates the overall trend in Terry Gross' fundamental frequency from 2000-2022. The results indicated that Terry Gross' fundamental frequency decreased as she aged ($R^2 = 0.004$, $F(1, 51307) = 210.6$, $p < 0.01$).

a) b)

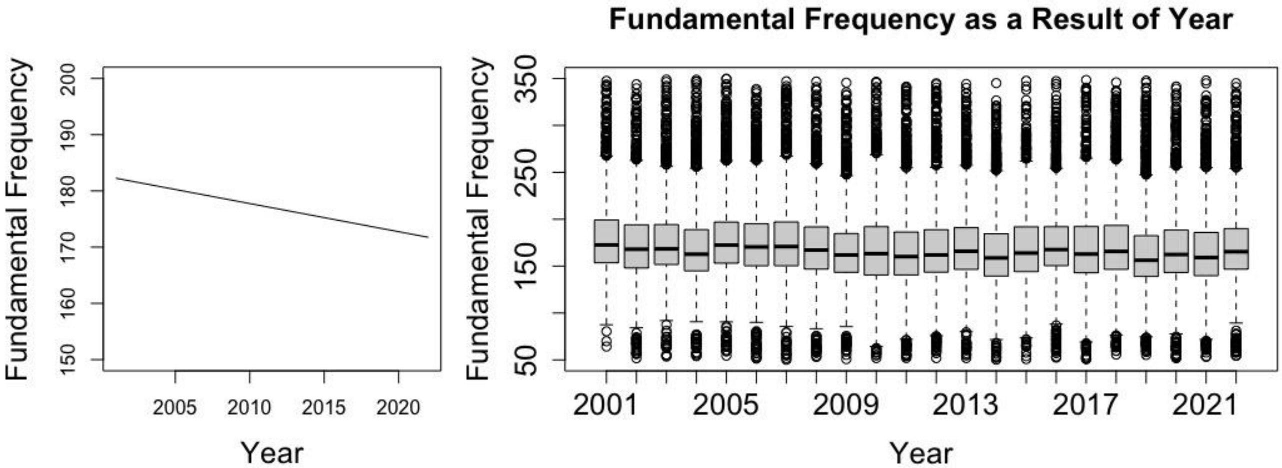


Figure 1: The modeled effect of year on fundamental frequency of Terry Gross over 21 years (a) and the linear regression showing the change in fundamental frequency over time (b).

The next variables that were analyzed were the first, second and third formant frequencies. Like fundamental frequency, for each of the formant frequencies, I ran separate linear regression models. Each linear regression model involved the formant frequency (whether it was 1^{st}, 2^{nd}, or 3^{rd}) as a function of year. The results indicated that as time passed, the value of the first formant frequency increases ($R^2 = 0.0047$, $F(1, 53751) = 26.28$, $p < 0.01$). As shown in Figure 2 below, where there is a slight increase in F1 with increasing age. Figure 2a demonstrates the modeled effect of years on F1, which demonstrates that over this timespan, Terry Gross' F1 increases by 7.16 Hz. Figure 2b is the linear regression of F1 plotted as a function of years.

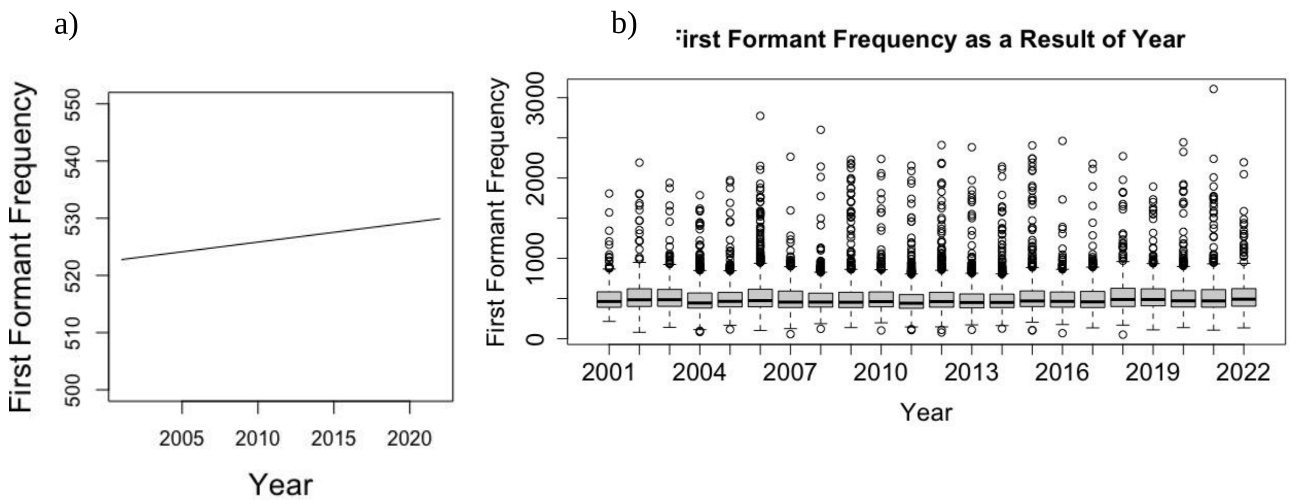


Figure 2: The modeled effect of year on the first formant frequency (a) and the linear regression demonstrating the shift in Terry Gross's F1 over 21 years (b).

The value of the second formant frequency showed dissimilar results to the first formant frequency and the fundamental frequency. For this acoustic feature, there was no significant effect of year on the value of the second formant frequency (R^2 = 2.406e-06, F (1, 53751) =

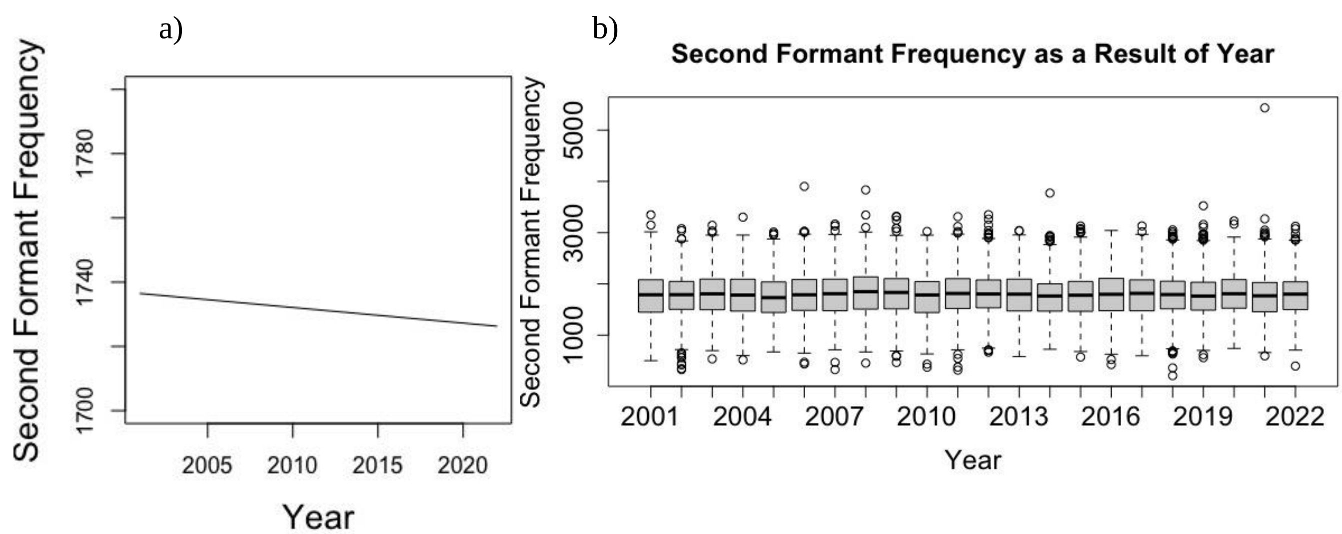


Figure 3: The modeled effect of year on F2 (a) and the linear regression demonstrating the shift in Terry Gross's F2 over 21 years (b).

1.129, $p < 0.2879$). Since the p value is greater than 0.05, it is not possible to reject the null hypothesis that year has no effect on second formant frequency. Based on Figure 3a above showing the modeled effect of year on F2, as well as Figure 3b, it appears as if F2 is decreasing slightly over the years. Although the effect of year on F2 is found to be a decrease in 10.19 Hz, this effect is not significant.

2.2.3 Results for the acoustic analysis of Ira Glass

As with Terry Gross, the first acoustic characteristic that I modeled as a function of year was the fundamental frequency of Ira Glass's voice. I posit that the results found in the analysis of Ira Glass' voice will be consistent with those found for Terry Gross. Upon modeling F0 as a function of year, the output demonstrated that as years increase, there is a decrease in Ira Glass' fundamental frequency ($R^2 = 0.0028$, $F(1, 109070) = 303.4$, $p < 0.01$). Figure 4a is a plot showing the modeled effect of year on the fundamental frequency of Ira Glass. It shows how the change in years has an effect of 5.33 Hz on the fundamental frequency over the course of Ira Glass' lifespan. Figure 4b is a linear regression which demonstrates the overall trend in Ira Glass' fundamental frequency from 2000-2022.

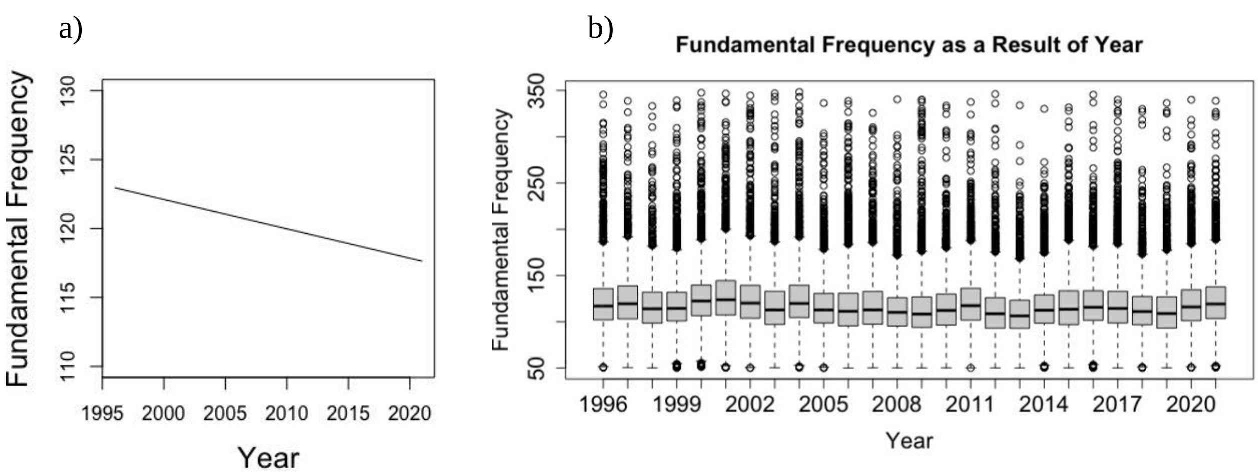


Figure 4: Modeled effect of year on fundamental frequency (a) and the linear regression demonstrating the change in fundamental frequency over the years (b).

The next variables that were analyzed were the first, second and third formant frequencies. Like the fundamental frequency, for each of the formant frequencies, a separate linear regression was run. Each linear regression model involved the formant frequency (whether it was 1st, 2nd, or 3rd) as a function of year. For F1, as time passed, F1 increased ($R^2 = 0.0001695$, $F(1, 121611) = 21.62$, $p < 0.01$). Figure 5a shows the modeled effect of year on the F1 of Ira Glass. According to this, as well as the linear regression shown in Figure 5b, the effect of year on F1 is an increase of 7.64 Hz over the timespan used.

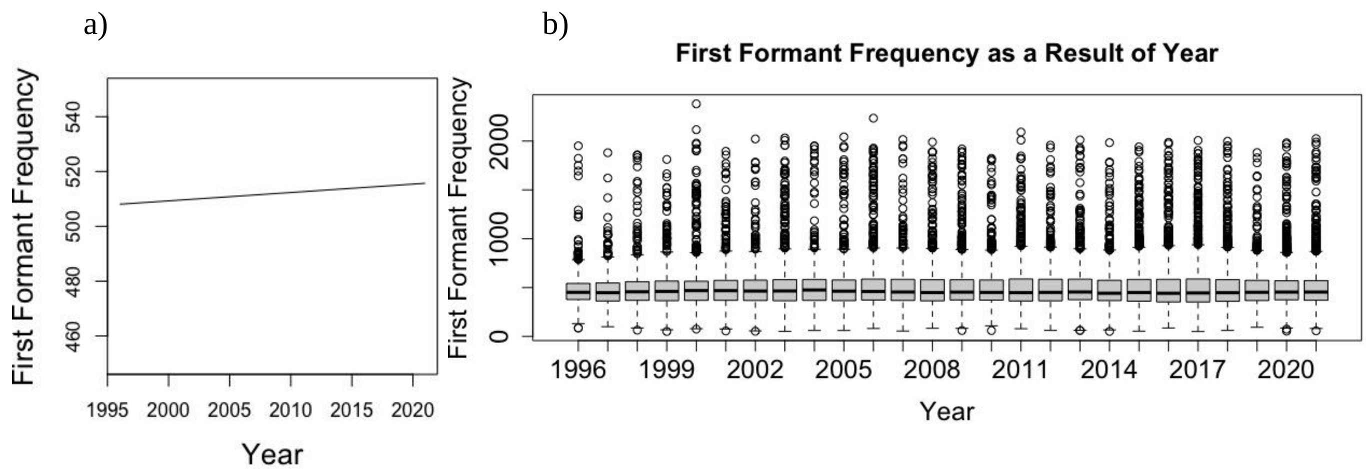


Figure 5: Modeled effect of years on the first formant frequency (a) and the linear regression demonstrating the change in F1 over the years (b).

The value of the second formant frequency showed similar results to the first formant frequency and the fundamental frequency, unlike the second formant frequency of Terry Gross. That is, Ira Glass' F2, like F1, increased as years passed (opposite to that mentioned by prior research) ($R^2 = 0.00006$, $F(1, 121611) = 8.3$, $p < 0.005$)**.** Figure 6a demonstrates the modeled effect of year on F2, which is an increase of 7.64 Hz. over Ira Glass' lifespan given. Figure 6b below shows the linear regression.

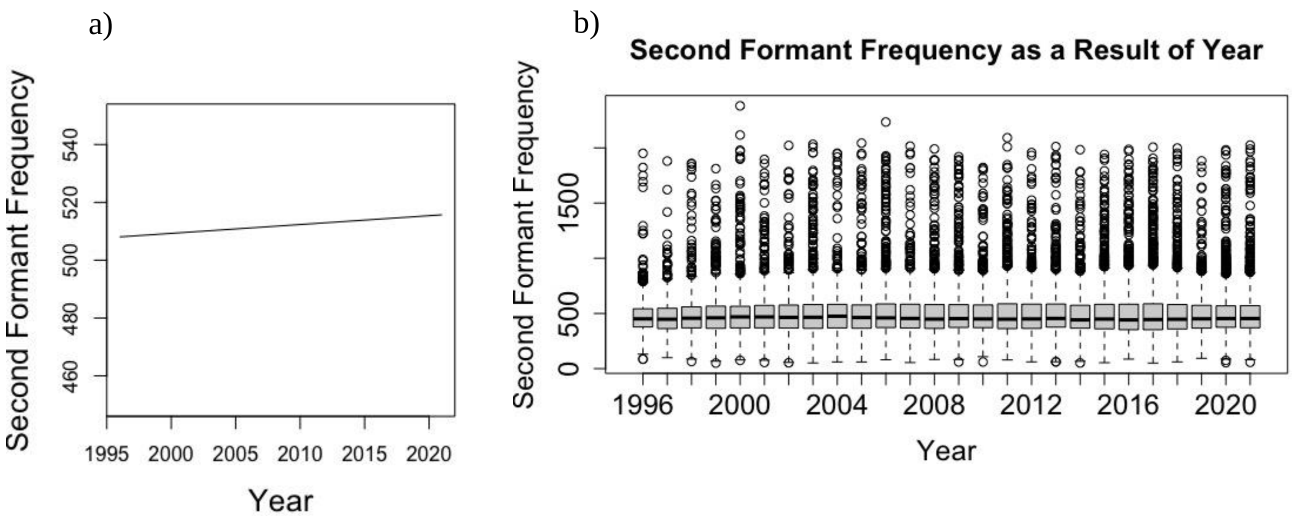


Figure 6: Modeled effect of year on second formant frequency (a) and the linear regression demonstrating the shift in Ira Glass's F2 over 25 years (b).

2.3 Discussion

To sum up the results of the above analysis, I have observed shifts in the acoustic characteristics of Terry Gross and Ira Glass over the course of both of their lifespans. For Terry Gross, her fundamental frequency and F1 increase. Terry Gross' F2 appeared to decrease; however, the result was not significant. For Ira Glass, on the other hand, fundamental frequency decreased, while his F1 and F2 increased.

Compared to the results of other studies, the results of the analysis performed on Terry Gross and Ira Glass both agreed and disagreed with some prior work. Regarding F0, the results of both Terry Gross and Ira Glass showed that as years increase, F0 decreases. This is in agreement with a lot of earlier work including that of Fisher and Linville (1985) and Reubold, Harrington, and Kleber (2010). Based on Figure 1b and 4b it appears that the fundamental frequencies of Terry Gross and Ira Glass may also be starting to follow the U shaped pattern mentioned by Reubold, Harrington, and Kleber (2010) in the later years of the recordings. However, to confirm this, a larger age range would be helpful. What can be confirmed based on Figure 1a and 4a is that F0 for both speakers are decreasing over the lifespan.

Regarding the formant frequencies of the vowels, we see that for F1, the work of Fisher and Linville (1985) found that F1 decreased with an increase in years, which is opposite to what I found. Linville and Fisher (1985) found that F2 decreases as age increases for women; however, for Terry Gross, the decrease was not found to be significant. This may be due to the smaller time span. For Ira Glass, F2 increased, which opposes the work conducted by Eichhorn et al. (2018), which shows a decrease in F2 in males over the lifespan.

CHAPTER 3. EXPERIMENT 1: TERRY GROSS

In the previous chapter I showed that the voices of our two speakers have changed with age. For Terry Gross, there was a decrease in her fundamental frequency, as well as an increase in her F1. While a decrease in her F2 was visible in Figure 3, the effect of years was not statistically significant. In this chapter, I investigate whether the changes in Terry Gross' acoustic characteristics over her lifespan are perceptually salient in an AxB comparison task.

3.1 Methodology

3.1.1 Participants

One hundred and sixty-seven participants, ranging from 17-28 years old, were recruited from the University of Alberta, Department of Linguistics subject pool, a subject pool that consists of undergraduate students enrolled in introductory linguistics classes. When asked about their gender in the demographic survey, of the 167 participants, 100 chose female, 58 chose male, 6 chose non-binary, and 3 participants preferred not to answer. There were also a range of first languages, including: English (n = 114), Mandarin (n = 22), Arabic (n = 5), Korean (n = 4), Tagalog (n = 4), Cantonese (n = 3), Hindi (n = 2), Japanese (n = 2), Vietnamese (n = 2), Urdu (n = 2), Romanian (n = 2), Spanish (n = 2), Russian (n = 1), French (n = 1), Bengali (n = 1). Listeners participated for less than 60 minutes and received 2 course credits as compensation for their participation.

3.1.2 Data

The data used in this experiment was Terry Gross' voice data from the acoustic analysis performed in Chapter 2, which included 22 recordings downloaded from National Public Radio (see Chapter 2).

3.1.3 Materials

For the purposes of this experiment, I used 22 recordings from January 2000 until January 2022. For the age perception experiment, I cut down each of the two recordings I took per year to be about three seconds long (the 3 seconds was taken from different parts in each recording, depending on if there was a short phrase that fit well in this timespan), so that the time of the experiment could be controlled. Originally, each recording downloaded from NPR was over an hour long, and for the purposes of this experiment, only a short speech sample was required from each. After editing the recordings down into 3 second clips, a custom Praat script was used to normalize the amplitude of each of the recordings (so that each of the 22 recordings have the same amplitude), in order to avoid any influence that different amplitudes may have had on the perception of age by the listeners.

All the data was collected using an online experiment that was designed using JavaScript. Experiment 1 using the voice of Terry Gross contained 462 comparisons, in which 2 recordings are compared (for example, a 2020 recording is compared with a 2001 recording). Each year was compared to the other 21 years and each comparison was repeated twice to check for consistency. At the beginning of the experiment a demographic survey asked for age, gender, and language information (see A.2 in Appendix A). At the end of the experiment there were also four Likert scale questions in which the participant was asked questions aimed to determine their interaction frequency (with individuals over 40), their interaction positivity (with people over 40), their enjoyment (of time spent with people over 40) and if they value the time they spend with people over 40 (see A.3 in Appendix A). Finally, a short-answer question prompting participants to identify any speech characteristics they heard in the speech samples that indicated to them which recording might contain the older voice (see A.4 in Appendix A).

3.1.4 Procedure

Upon receiving ethical approval from the University of Alberta Research Ethics Board (Pro00124229), the experiment, designed using JavaScript, was posted to the Department of Linguistics subject pool along with a description so that students within the pool would be able to sign up. After students chose and signed up for the experiment titled "Guess My Age," they received a link to the experiment website, which began with a consent form. Participants could not move forward before consenting. The study was described as one in which the participant, listens to two audio samples and determines in which one of the 2 recordings the speaker's voice sounds older. After consenting to the experiment, participants filled out information regarding their gender, age, native language, and languages spoken. Following this, the experiment moves on to the first comparison. In each comparison, the participant will listen to Stimulus A, followed by a short, silent break, then Stimulus B. After listening to both, the participant was prompted with the question "Does the speaker sound older in Sound A or Sound B," and asked to click A or B on the keyboard depending on which they thought sounded older. Upon completion of the 462 comparisons, the participant moved on to the Likert scale questions and the short-answer question asking about the acoustic features they had relied on during the age perception task.

Participants received the comparisons in different orders; however, each participant received all 462 comparisons. Participants completed the experiment online, remotely, and independently, prior to the deadline. The participants were asked to complete the study on computers (the study was not compatible with mobile devices), in a quiet, distraction-free environment while wearing headphones. The study was designed to be done in one sitting, taking approximately one hour. At the end of the experiment students received a thank-you screen with instructions to receive credit, as well as the debriefing form.

3.2 Results

Data analysis was conducted using generalized linear-mixed effects models in R (R Core Team 2022) using the lme4 package (Bates et al. 2015), in order to estimate the effects of random and fixed effects on the dependent variable, accuracy. Generalized models were chosen in this study as the dependent variable is binary (the answer given by the participant was either A or B, which was either correct or incorrect) and not continuous.

All generalized linear-mixed effects models created to visualize the listener data for Terry Gross involved using the dependent variable of accuracy, which is whether the speaker determined the correct recordings as having the older voice as the outcome. In total, 9 models were created to determine which variables acted as predictors and had significant effects on the accuracy. The predictors that were investigated were: difference (in years between recordings), age, sex, native language, interaction frequency (with individuals over 40), interaction positivity (of time with people over 40), enjoyment (of time spent with people over 40) and if they valued the time they spend with people over 40. Subject was kept in each model as a random effect. In all the models, including variables other than difference did not improve the fit of the model. Upon testing a model where there was an interaction between native language and difference it was determined that there is a trending effect that non-native English speakers have poorer accuracy; however, this effect is not significant ($p > 0.05$). The final model included difference of years between recordings as a predictor, while keeping subject as a random effect, and accuracy of a listener as the dependent variable. In this model, difference in years between recordings had a significant effect on accuracy of a listener (beta = 0.015, $t(167) = 10.27$, $p < .001$), specifically it had a positive significant effect on the accuracy of a listener. It is important to note that a warning message did appear in the summary for this model, suggesting a rescaling of the variable difference in years between recordings. AIC tests, which are tests used to determine goodness of fit of a model, were used to compare models, and the final model using difference as the only predictor had the fewest degrees of freedom as well as the best model fit, and was thus chosen due to its simplicity, as well as significance.

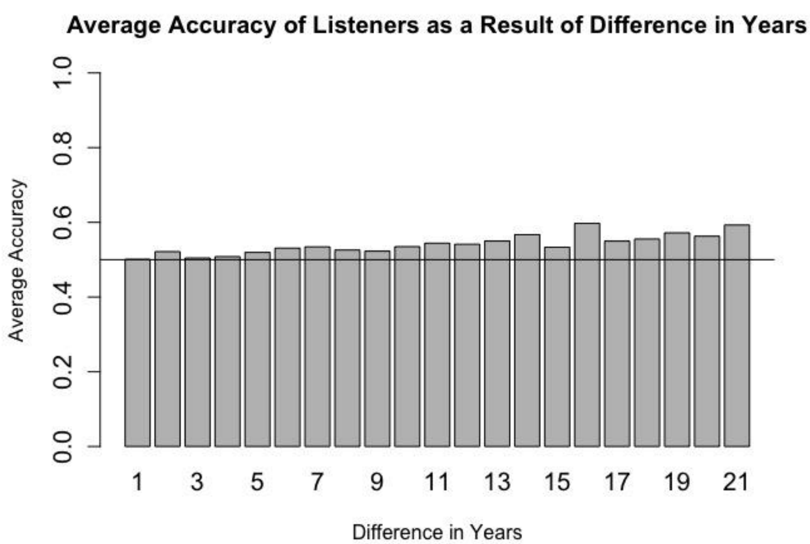


Figure 7: Accuracy of listeners based on the difference of years for Terry Gross with an intersection showing chance.

Figure 7 demonstrates how as the difference in years increases, the accuracy of the speaker increases as well. It appears that by 21 years difference in recordings, speakers are slightly more accurate. For every unit increase in year, there is an increase of 1.5% in accuracy.

In Figure 7, the ab line intersects at 0.5, or at chance. Based on the bar plot one can see that listeners are better than chance at determining which recording is of an older Terry Gross, with an accuracy of 62% for a difference of 21 years**.**

On the short answer questions participants were asked to answer at the end of the experiment, they demonstrated strong opinions on what they thought constituted older speech. Overall, individuals said that stuttering, a slowed speech rate, decreased pitch and a raspy voice quality in the voice distinguished an older voice from a younger voice.

3.3 Discussion

Based on the results of Experiment 1, it is observed that as the years between recordings increases, so does the accuracy. The results are not the results expected, as the accuracy is lower than anticipated based on previous studies with smaller gaps between recordings. Bruckl (2007) demonstrated a listener accuracy of about 70% ("listeners were accurate 2/3 of the time" (Bruckl 2007, 3)) with a 5-year gap between recordings; however, the accuracy of listeners in Experiment 1 is at about 60%, or just above chance, with a gap of 21 years between the youngest and oldest speech samples. Based on these results, I hypothesize that an even larger gap between recordings will improve the accuracy of listeners.

CHAPTER 4. EXPERIMENT 2: IRA GLASS

In the previous chapter I showed that listeners are slightly above chance in discriminating the older voice out of two recordings when there is a span of 22 years. In other words, the results showed that listeners had the highest accuracy when there was the largest possible difference between recordings (which in the case of Terry Gross, was 22 years). These results encouraged me to reattempt the AxB comparison task; however, this time with a larger time span, that being the 25 years of speech of Ira Glass I have available. Ira Glass' acoustic characteristics shifted during his lifespan, as suggested by the results of the acoustic analysis conducted in Chapter 2. Thus, this chapter, aims to determine not only whether Ira Glass' shifting acoustic characteristics are perceptually salient during an AxB task, but also whether listeners' accuracy in discriminating the older voice increases compared to that of listeners in Experiment 1, due to the increased number of years of data available.

4.1 Methodology

4.1.1 Participants

Ninety-four participants were recruited from the Department of Linguistics subject pool. When asked on the demographic survey about gender, 59 participants chose female, 30 chose male, 4 chose non-binary, and one preferred not to answer. The youngest participant was 17, while the oldest was 45. The native languages of the participants included: English (n = 60), Mandarin (n = 8), Japanese (n = 3), Russian (n = 3), Korean (n = 3), Somalian (n = 2), Tagalog (n = 2), Arabic (n = 2), Punjabi (n = 2), Hindi (n = 1), Indonesian (n = 1), Italian (n = 1), Turkish (n = 1), Tamil (n = 1), Spanish (n = 1), Urdu (n = 1), Bengali (n = 1) and one individual who preferred not to answer. Participants received 3% course credit for Experiment 2 as compensation for their participation.

4.1.2 Data

The data used in Experiment 2 was Ira Glass' voice data from the acoustic analysis performed in Chapter 2, which included 25 recordings downloaded from National Public Radio (see Chapter 2, section 2.1 for description of the stimulus source).

4.1.3 Materials

For Ira Glass, I took one recording from each year from 1996 to 2021. Specifically, I took the recording from January of each year. His recordings and transcripts were all downloaded from the NPR archives (from his show, "This American Life"). For the purposes of this sub project, I used 25 of the recordings from January 1996 until January 2021. For the age perception experiment, I cut down each of the recordings I took per year to be about three seconds long, so that the time of the experiment could be controlled. Since there were more recordings for Ira Glass (since he had a longer timespan), the recordings had to be slightly shorter than those used in Experiment 1. After editing the recordings down into 3 second clips, a custom Praat script was used to normalize the amplitude of each of the recordings (so that each of the 25 recordings have the same amplitude), as done in Experiment 1 in Chapter 3.

All the data was collected using an online experiment that was designed using JavaScript. Experiment 2 follows an identical procedure to Experiment 1 (see section 3.1.4 for details) but uses stimuli from Ira Glass rather than Terry Gross. Experiment 2 using the voice of Ira Glass contained 553 comparisons in which 2 recordings are being compared (for example, 2020 is being compared with 2001). Each of the 25 years was compared to the other 24 years and each comparison was repeated twice, to check for consistency (this is where 553 comparisons come from). At the beginning of the experiment a survey asked for age, gender, and language information (see A.2 in Appendix A). At the end of the experiment the same four Likert scale questions as those used in Experiment 1 were presented to participants (see A.3 in Appendix A). Finally, like in Experiment 1, a short-answer question was given prompting participants to identify any acoustic features they relied on to discriminate the older voice (see A.4 in Appendix A).

4.1.4 Procedure

The procedure followed for Experiment 2 was identical to that of Experiment 1. The main difference between the two experiments was that Experiment 2 contained 553 comparisons, rather than 462 like in Experiment 1. This increase is due to the increase in years available in Experiment 2 and resulted in Experiment 2 being slightly longer than Experiment 1, which is why participants received an extra credit as compensation for participation in Experiment 2.

4.2 Results

Like for Experiment 1, data analysis was conducted using generalized linear-mixed effects models in R (R Core Team 2022) using the lme4 package (Bates et al. 2015), in order to estimate the effects of random and fixed effects on the dependent variable. As with Terry Gross, modelling was used to determine the effects of a range of variables (difference of years between recordings, age, sex, native language, interaction frequency with people over 40, interaction positivity, enjoyment regarding time spend with people over 40, and if participants valued time spent with people over 40) on the accuracy of a given listener in the discrimination task. A similar process to that used in Experiment 1 was used to determine the model with the best fit.

All generalized linear-mixed effects models created to visualize the listener produced data for Ira Glass involved using the dependent variable of accuracy, which is whether the speaker determined the correct recordings as having the older voice as the outcome. In total, 5 models were created to determine which variables acted as predictors and had significant effects on the accuracy. The predictors that were investigated were: difference (in years between recordings), age, sex, native language, interaction frequency (with individuals over 40), interaction positivity, enjoyment (of time spent with people over 40) and whether participants value the time they spend with people over 40. Subject was kept in each model as a random effect. In all the models, including variables other than difference did not improve the fit of the model. The final model included difference of years between recordings as a predictor, while keeping subject as a random effect, and accuracy of a listener as the outcome. In this model, the main effect of difference on accuracy of a listener was not significant (beta = -.008, $t(97) = -0.52$, $p > .05$). It is important to note that a warning message appeared suggesting that the dependent variable be rescaled. After each model was created, AIC tests were run to determine which model had the best fit. Through the process, it was determined that the model in which difference between years is the only predictor for accuracy had the fewest degrees of freedom as well as the best fit. Although this model is the best fit model, none of the models had variables that had a significant effect on the accuracy of a listener.

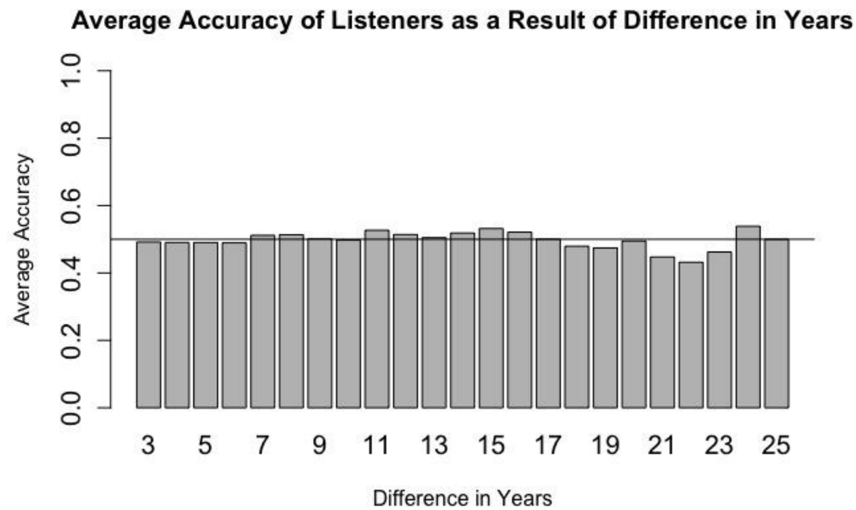


Figure 8: Accuracy of listeners based on the difference of years for Ira Glass.

Figure 8 illustrates how as the difference in years increases, the accuracy of the speaker seems like it could be increasing; however, then it dips down after a 15-year difference. Like in Figure 7, Figure 8 has a horizontal line that intersects at 0.5, or at chance. Based on the bar plot one can see that listeners are at chance, or even worse, at discriminating age in Ira Glass recordings. In other words, it appears as if listeners are guessing. Even though I found some acoustic differences in Ira Glass' voice in the acoustic analysis (see Chapter 2), it was not enough for listeners to perceive difference in age.

On the short answer questions participants were asked to answer at the end of the experiment, they demonstrated strong opinions on what they thought constituted older speech. Overall, individuals said that a slowed speech rate, decreased pitch, a raspy voice quality and lower amplitude characterized an older voice. The opinions shown by listeners on Experiment 2's short answer question are nearly identical to those on Experiment 1 (see Chapter 3).

4.3 Discussion

As with the results of Experiment 1, the results of Experiment 2 contrasted my original hypothesis. Initially, I hypothesized that an increase in years between recordings would increase the accuracy of a listener, opposite to what the results have shown. The results of Experiment 2 showed that an increase in years between recordings did not improve accuracy. In Experiment 2 listeners' accuracy was at chance (with a maximum difference of 25 years between recordings), which is lower than the accuracy found in Experiment 1(which had a maximum of 22 years difference between recordings). Since I increased the gap between the recordings from 21 (Experiment 1) to 25 (Experiment 2), I anticipated an increase in accuracy along with the increase in years. However, accuracy decreased to about chance. The explanation for this could be related to the age of Ira Glass in the recordings. The oldest age Ira Glass is in the recordings is 63, where Terry Gross is 71 at her oldest. A previous study by Ferguson (2013) in which a group of listeners (17-50 years old) were asked to try to categorize a voice's age into one of three age categories (50-66, 67-83, 84-100) showed that listeners are more accurate at categorization if a speaker is older than 68. This research demonstrates that there is a chance that since Ira Glass's recordings do not make it to that threshold number of 68, his voice has not changed enough for individuals to be able to discriminate between his older voice and his younger voice.

CHAPTER 5. GENERAL DISCUSSION & CONCLUSION

5.1 General discussion

In the previous chapters I presented an acoustic analysis, as well as two perceptual experiments (involving the voices of Terry Gross and Ira Glass), which aimed to determine how the voices of two speakers change with age, as well as whether speakers can discriminate the age of a speaker between two recordings that are from different years. Although in the acoustic analysis performed in Chapter 2 I found that the acoustic characteristics (specifically, F0, F1 and F2) of Terry Gross' and Ira Glass' voices **do** change over the course of their lifespans, these changes were not perceptually salient in the case of Experiment 2. The best results achieved were during Experiment 1, in which participants listened to the comparison of one recording of Terry Gross from 2000 and one from 2022. It was found that listeners were slightly above chance when identifying the correct recording as being from an older Terry Gross. There are two possible explanations as to why listeners are slightly better at discriminating between the two recordings of Terry Gross to determine which recording is from when she is older. The first explanation, discussed in the previous section, may be the age threshold. For Terry Gross, her age ceiling is 71, where the Ira Glass recordings used here stop when he is 63. A previous study by Ferguson (2013) in which a group of listeners (17-50 years old) were asked to try to categorize a voices age into one of three age categories (50-66, 67-83, 84-100) showed that not only are listeners better at discriminating when there is a) a greater gap in years between the recordings, but b) listeners are more accurate at categorization if a speaker is older than 68. This research could explain why participants have a higher accuracy for discrimination with Terry Gross. Since her oldest recording involves her voice at the age of 71, it is above the threshold arrived at by

Ferguson et al. (2013), and thus may support the validity of a threshold age for increased accuracy of discrimination.

Another possible explanation for the low accuracy could be that humans are simply bad at these discrimination tasks, even though they are confident in their ability on the task. For example, participants had clear ideas of what they thought constituted older speech. During the short answer question given at the end of Experiments 1 and 2, participants expressed that they chose the 'stuttering, slowed speech" or that they "tend to associate older voices with more slurred-sounding speech, or with a kind of raspy quality to them." Another participant even said, "slowed voice indicated older to me, while a faster voice indicated younger." These responses demonstrate that individuals have preconceived assumptions of what older speech sounds like. These assumptions, which occurred frequently, demonstrated participants beliefs and consequently, their inaccuracy. Using voice to estimate age is not the only subconscious judgements listeners make. Research performed by Barreda (2016) showed that listeners have confidence in their capability of making other assumptions of people based on their voice, such as estimations of a person's size; however, the results of this study showed that people were not accurate in their estimates (Barreda 2016, 15). This type of research demonstrates the type of judgements we make about people based on certain stereotypical characteristics.

5.1.1 Future directions

Future research should focus on acquiring long-term data from a speaker's lifespan. For example, should Ira Glass continue to host "This American Life," the acoustic analysis and perceptual experiment using Ira Glass' voice performed in this thesis could be replicated when Ira Glass is older and has more speech available. This could help us to identify if an increased difference in years between recordings could improve a listener's accuracy in the discrimination task. Other avenues of future research involve looking deeper into the sociological implications of "older" speech characteristics; for example, are the notions of "slowed, stuttering speech" I have seen associated with older speech in this experiment consistent culturally? Additionally, I could also investigate how individuals alter their speech based on these stereotypical markers of "older" speech; for example, if a participant listened to an interrogative speech sample with certain characteristics they may have opinions on, see how they respond, and ask what characteristics they are specifically responding to.

5.2 Conclusion from the acoustic analysis

The goal of this project was to determine if voices change with age, and more specifically to the data collected, if the voices of Terry Gross and Ira Glass change over certain periods of their individual lifespans. Based on the results, it can be determined that certain acoustic characteristics of voice such as fundamental frequency (F0), and the first, second and third formant frequencies change with age. Furthermore, the results suggest that there was also a difference between how characteristics shift between male and female speakers. Although fundamental frequency decreases with time for both males and females, the formant frequencies of males seem to increase over time, while the formant frequencies of females decrease over time. So, male and female voices shift to more of a median value over time (each going opposite directions). A greater number of speakers that share similar characteristics to those used in this project would be needed to confirm these results, and hopefully, an increase in good quality, clear audio recordings will be archived for purposes like these. Studies like this also have the potential to aid in the synthesis of aged voices, and to further investigate stereotypes related to aging.

5.3 Conclusion of the perception experiments with Terry Gross and Ira Glass

The results of this thesis suggest that individuals listening to two recordings of the same speaker from two different years have an above chance possibility of determining which recording is older if the speaker's oldest age is above 68. If the age does not meet this threshold, like Experiment 2 using the voice of Ira Glass, listeners have a 50% chance of discrimination. The characteristics of the listeners such as age, sex, native language, frequency/positivity/enjoyment/value of time with people over 40 do not seem to have any significant effects on the accuracy of discrimination. Although I originally thought that factors such as individuals' interaction frequency, positivity or enjoyment of time spent with people over 40 would have a significant impact on their accuracy during the perception experiments, it did not. My reason for believing this was because I thought that people that spent more time with people over 40 would have more chances to hear and identify speech from this older group.

5.4 Final thoughts

This research has highlighted the fact that we tend to make subconscious assumptions about individuals based on their voice. I believe this research demonstrates potentially harmful preconceived notions we may have of what "older speech" sounds like. This research makes it clear that just because we hear a feature such as slower speech rate, or a stutter, does not indicate an older speaker. It is important that we ignore these assumptions because they can impact how we choose to use our voice in a conversation with someone, which can have negative effects on our conversation partners. Further research should aim to gather speech samples from speakers that span over 20 years and have speech from an individual's lifespan from when they are very young to very old (old being closer to 100 than 70) to determine if span, or age range of the speaker is the variable which allows for increased accuracy of discrimination by a speaker.

REFERENCES

Barreda, Santiago. 2016. "Investigating the Use of Formant Frequencies in Listener Judgments of Speaker Size." *Journal of Phonetics* 55 (March): 1–18. https://doi.org/10.1016/j.wocn.2015.11.004.

Bates, Douglas, Martin Mächler, Ben Bolker, and Steve Walker. 2015. "Fitting Linear Mixed-Effects Models Using Lme4." *Journal of Statistical Software* 67 (1). https://doi.org/10.18637/jss.v067.i01.

Brown, W.S., Richard J. Morris, Harry Hollien, and Elizabeth Howell. 1991. "Speaking Fundamental Frequency Characteristics as a Function of Age and Professional Singing." *Journal of Voice* 5 (4): 310–15. https://doi.org/10.1016/S0892-1997(05)80061-X.

Bruckl, Markus. 2007. "Women's Vocal Aging: A Longitudinal Approach." *INTERSPEECH*, 1170–73.

Eichhorn, Julie Traub, Raymond D. Kent, Diane Austin, and Houri K. Vorperian. 2018. "Effects of Aging on Vocal Fundamental Frequency and Vowel Formants in Men and Women." *Journal of Voice* 32 (5): 644.e1-644.e9. https://doi.org/10.1016/j.jvoice.2017.08.003.

Ferguson, Sarah H., Eric J. Hunter, Catherine A. Mellum, and Lydia R. Rogers. 2013. "Listener Judgments of Age in a Single-Talker 48-Year Longitudinal Sample." In , 060079–060079. Montreal, Canada. https://doi.org/10.1121/1.4799012.

Fisher, Hilda, and Sue Ellen Linville. 1985. "Acoustic Characteristics of Women's Voices with Advancing Age." *Journal of Gerontology* 40 (3): 324–30. https://doi.org/10.1093/geronj/40.3.324.

Harnsberger, James D., Rahul Shrivastav, William Brown, Howard Rothman, and Harry Hollien. 2008. “Speaking Rate and Fundamental Frequency as Speech Cues to Perceived Age.” *Journal of Voice* 22 (1): 58–69. https://doi.org/10.1016/j.jvoice.2006.07.004.

Hunter, Eric J., Sarah Hargus Ferguson, and Catherine Anne Newman. 2016. “Listener Estimations of Talker Age: A Meta-Analysis of the Literature.” *Logopedics Phoniatrics Vocology* 41 (3): 101–5. https://doi.org/10.3109/14015439.2015.1009160.

Hunter, Eric J., Mara Kapsner-Smith, Patrick Pead, Megan Z. Engar, and Wesley R. Brown. 2012. “Age and Speech Production: A 50-Year Longitudinal Study.” *Journal of the American Geriatrics Society* 60 (6): 1175–77. https://doi.org/10.1111/j.1532-5415.2012.03983.x.

Huntley, Ruth, Harry Hollien, and Thomas Shipp. 1987. “Influences of Listener Characteristics on Perceived Age Estimations.” *Journal of Voice* 1 (1): 49–52. https://doi.org/10.1016/S0892-1997(87)80024-3.

Klautau, Aldebaro. 2001. “ARPABET and the TIMIT Alphabet.” https://web.archive.org/web/20160603180727/http://www.laps.ufpa.br/aldebaro/papers/ak_arpabet01.pdf.

Linville, Sue Ellen. 1995. “Vocal Aging.” *Current Opinion in Otolaryngology & Head and Neck Surgery* 3 (3): 183–87.

Moran, Michael J., Laurie McCloskey, and Barbara Cady. 1995. “Listener Age Estimates of Elderly African American and Caucasian Male Speakers.” *Journal of Cross-Cultural Psychology* 26 (6): 751–58. https://doi.org/10.1177/002202219502600614.

R Core Team. 2022. “R: A Language and Environment for Statistical Computing.” Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Reubold, Ulrich, Jonathan Harrington, and Felicitas Kleber. 2010. “Vocal Aging Effects on F0 and the First Formant: A Longitudinal Analysis in Adult Speakers.” *Speech Communication* 52 (7–8): 638–51. https://doi.org/10.1016/j.specom.2010.02.012.

Ryan, Ellen, and Harry Capadano. 1978. “Age Perceptions and Evaluative Reactions Toward Adult Speakers.” *Journal of Gerontology* 33 (1): 98–102. https://doi.org/10.1093/geronj/33.1.98.

Ryan, William, and Kenneth Burk. 1974. “Perceptual and Acoustic Correlates of Aging in the Speech of Males.” *Journal of Communication Disorders* 7 (2): 181–92. https://doi.org/10.1016/0021-9924(74)90030-6.

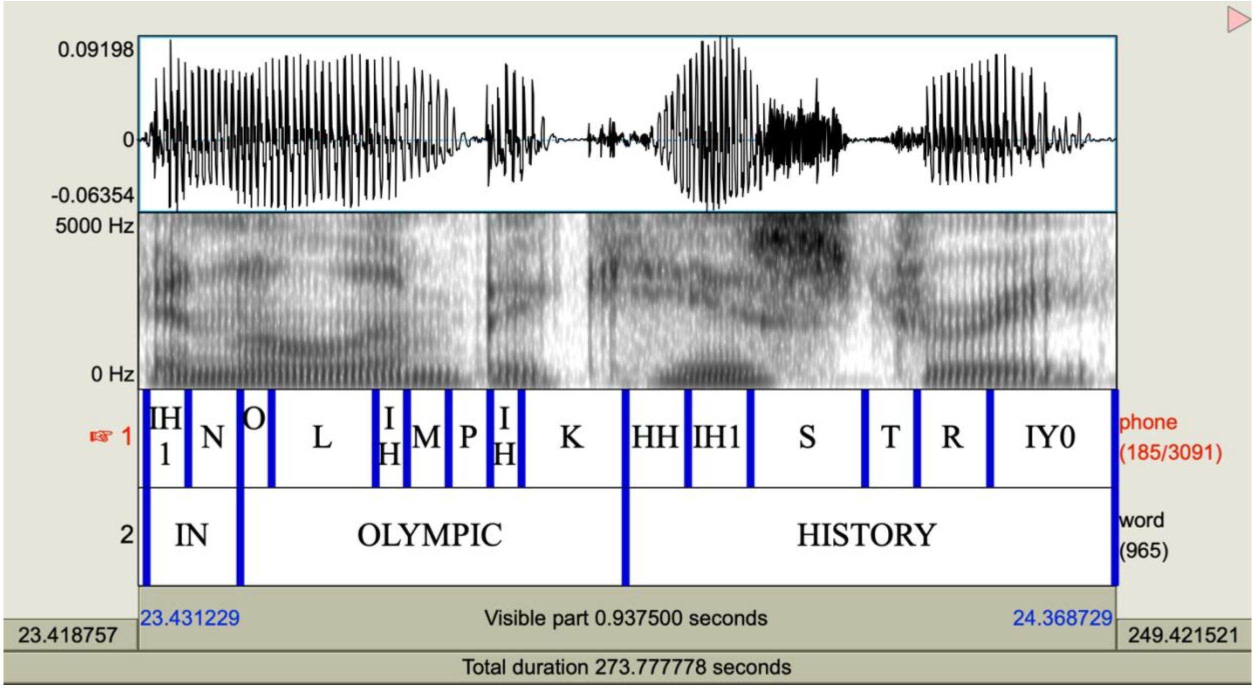
Stathopoulos, Elaine T., Jessica E. Huber, and Joan E. Sussman. 2011. “Changes in Acoustic Characteristics of the Voice Across the Life Span: Measures From Individuals 4–93 Years of Age.” *Journal of Speech, Language, and Hearing Research* 54 (4): 1011–21. https://doi.org/10.1044/1092-4388(2010/10-0036).

Tucker, Benjamin V., Catherine Ford, and Stephanie Hedges. 2021. “Speech Aging: Production and Perception.” *WIREs Cognitive Science* 12 (5). https://doi.org/10.1002/wcs.1557.

Yuan, Jiahong, and Mark Liberman. 2008. “Speaker Identification on the SCOTUS Corpus.” Proceedings of Acoustics ’08.

APPENDIX A

A.1: Examples Forced Alignment



A.2 A short answer demographic survey presented at the beginning of both Experiment 1 and 2

1) How old are you?
    a. Short answer
2) What is your native language?
    a. Short answer
3) What gender do you identify with?
    a. Short answer

A.3 Likert scale questions

1) How often do you interact with individuals over the age of 40?
    a. 7-point Likert scale, 1 = never, 7 = very often
2) How positive would you say that your interactions with people over the age of 40 are?
    a. 7-point Likert scale, 1 = very negative, 7 = very positive
3) How strongly do you agree with this statement: I enjoy the time I spend with individuals over 40.
    a. 7-point Likert scale, 1 = strongly disagree, 7 = strongly agree
4) How strongly do you agree with this statement: I value the time I spend with individuals over 40.
    a. 7-point Likert scale, 1 = strongly disagree, 7 = strongly agree

A.4 Fill in the blank question regarding specific vocal characteristics they used as indicators of certain speech age.

1) What acoustic characteristics did you use to help you determine the older speaker?
    a. Short answer