Technische Universität Berlin

Faculty IV- Electrical Engineering and Computer Science

Quality and Usability Lab

Perception of social speaker characteristics in natural speech

A thesis presents for the degree of Masters in computer science

January,2021

By: Mutasim Adam Mohamed

Supervisor: Prof. Dr-Ing. Sebastian Möller
Assistant Supervisor: Saisirisha Rallabandi

Declaration

I hereby declare that the thesis submitted is my own, unaided work, completed without any unpermitted external help. Only the sources and resources listed were used. The independent and unaided completion of the thesis is affirmed by affidavit.

Berlin, January 15,2021

_____

Mutasim Adam Mohamed

Dedications

I would like to dedicate this work to my parents who gave me my inspiration and my motivation. To my brother,sisters,friends, and classmate who participated in the survey study.

Acknowledgement

I would like to thank my supervisor Prof.Dr-Ing. Sebastian Möller. Also I wanted to thank the Assistant supervisor: Saisirisha Rallabandi ,for her supporting, giving technical issues, and advice to accomplish this study, I am grateful to have you as my supervisor,I very much appreciate your help.

Contents

List of Figures

List of Tables

0.1 Abstract

Speech uses to communicate with others, gain knowledge, and express our idea; thus, recognizing the speaker's characteristics became an important research topic in current social studies and motivated researchers to analyze the speech to recognize the speakers' cognitive aspects.

The judgment of a speaker's characteristics influences our social contact and affects numerous decisions that we make. This study will illustrate how a person's characteristics existing in the speaker's speech signal could be perceived as independent of speech utterance speakers. This work was accomplished by assigning a given sample of natural speech to women and men group, with two different scenarios, speech, and conversation, based on self-evaluative ratings. So those application domains like health care and customer service groups might benefit from Judgments of social characteristics.

In this experiment, 39 of the participants were involved in rating 52 attributes with a 5- point response scale. The tests were then conducted on the collected subjective ratings using statistical software packages for the social sciences (IMB SPSS). The results indicate that the social characteristics of the speaker could be perceived based on subjective ratings.

It also predicted a statistically significant difference between the female group and the male group in estimating the male speech, this can be noticed when we compare the mean of a female group (0.94) with the a mean of the male group (0.85). Simultaneously, there was no significant difference between female (mean =0.85) and male male group (mean= 0.90) groups in rating the female speech female.

Throughout this work, two core personality factors could be perceived from the speaker's voices.

0.2 Zusammenfassung

Sprache dient dazu mit anderen zu kommunizieren, Wissen zu erwerben und unsere Ideen auszudrücken. So wurde das Erkennen der Eigenschaften des Sprechers zu einem wichtigen Forschungsthema in aktuellen Sozialstudien und motivierte die Forscher, die Sprache zu analysieren, um die kognitiven Aspekte der Sprecher zu erkennen. Die Beurteilung der Eigenschaften eines Sprechers beeinflusst unseren sozialen Kontakt und auch zahlreiche Entscheidungen, die wir treffen. .

Diese Studie wird veranschaulichen wie die Eigenschaften einer Person, die in seiner Sprache vorhanden sind, als unabhängig von Sprachäußerungen betrachtet werden können. Die Arbeit wurde durchgeführt, indem einer Gruppe von Frauen und Männern eine bestimmte Stichprobe natürlicher Sprache zugewiesen wurde, mit zwei unterschiedlichen Szenarien, Sprachen und Konversationen, basierend auf Selbstbewertung. Daher könnten Anwendungsbereiche wie das Gesundheitswesen und Kundendienstgruppen von der Beurteilung sozialer Eigenschaften profitieren. In diesem Experiment waren 39 der Teilnehmer an der Bewertung von 52 Attributen mit einer 5-Punkte-Antwortskala beteiligt. .

Die Tests wurden dann mit Hilfe statistischer Softwarepakete für die Sozialwissenschaften (IMB SPSS) an den gesammelten subjektiven Bewertungen durchgeführt. Die Ergebnisse zeigen, dass die sozialen Eigenschaften des Sprechers anhand subjektiver Bewertungen wahrgenommen werden könnten. Es stellte sich auch ein statistisch signifikanter Unterschied zwischen der weiblichen und der männlichen Gruppe bei der Schätzung der männlichen Sprache heraus, dies wird deutlich, wenn der Mittelwert der weiblichen Gruppe (0,94) mit dem Mittelwert der männlichen Gruppe (0,85)

verglichen wird. Gleichzeitig gab es keinen signifikanten Unterschied zwischen der weiblichen und der männlichen Gruppe bei der Bewertung der weiblichen Sprache. Der weibliche Mittelwert liegt hier bei 0,85 und der männliche bei 0,90. Während dieser Arbeit konnten zwei zentrale Persönlichkeitsfaktoren anhand der Stimmen des Sprechers wahrgenommen werden.

Chapter 1

Introduction

1.1 Motivation

Speech is a communication method for humans[8].It is crucial to identify human personality, a specific type of speech denotes the specific type of speaker and displays speech traits. Thus, many recognition systems used a speech as an essential tool to identify the speaker's identity. Accordingly, numerous studies try to interpret and recognize speech information. Personality judgments play a significant role in our life in different ways, like decision planning, mate choosing, and job selection [14].

Moreover, the challenge goes beyond recognizing speech to analyze the speech and extract a piece of crucial information. Different traits have been widely addressed in many previous research, however, during a review of the literature related to this research area, it turned out that many researchers have evaluated different characteristics of the speaker by several methodological with synthetic speech, but a few modern researchers have judged a few social characteristics for natural speech dataset[19].

1.2 Overview of issues

Speaker Personality shows observable differences between humans, people's characteristics, and behavior. Diener & Lucas (2019) state that "Personality traits imply

consistency and stability—someone who scores high on a specific trait like extraversion is expected to be sociable in different situations and over time." [7].
According to Schuller et al. (2015), "The key-idea of personality psychology is that there are stable individual characteristics that explain most observable differences between people, especially when it comes to overt behavior,p.3"[20].

When we meet a person, it takes a short time to judge some personality, and form a first impression for some characteristics, like age, language, and gender, and make snap judgments and decisions. Nevertheless, this can lead to misjudgments because a person's perception can be affected by several variables and factors; thus, speech analysis came to address this problem. Since the speech is saturated with cues to a person's perceptions, the human voice is one of the main aspects to evaluate the speaker's identity. In this work, we conceptualized how to extract two fundamental dimensional of social perception in spontaneous speech and identify some personality traits. In other words, we suggested that there is a significant relationship between speaker characteristics and spontaneous speech. Speech is varying from spontaneous speech and synthesis speech. In this study, our case study is a natural speech as our dataset to judge the speaker's social characteristics.

1.3 Problem Definition and Research Question

Initial attempts focused on knowing the speaker's characteristics based on speech cues. This study's research goal is analyzing and perceiving some social characteristics of the speaker based on subjective evaluation scales and self-assessments. This work mainly tries to answer the question of how we could estimate the speaker's characteristics from his speech, or there is any significant relationship between speaker characteristics and his voice.

1.4 Structure Overview

This thesis is structured as follows:
In chapter 2, we highlighted the literature review and related approaches to this subject area. The overview of research design, research questions and hypothesis testing, and conducting subjective evaluation are discussed in chapter 3. Then, in chapter 4, The results and discussion are presented. Finally, in chapter 5, the work

summary and conclusion are covered.

Chapter 2

Literature Review

2.1 Theoretical Background

Human communication research shows that speech plays a vital role in both the recognition and perception of personality. Thus, speech analytic became very useful for customer domains and social science. It provides them with important information and helps many business sectors identify specific patterns and models to improve their services, automate many processes done manually, and improve speech technology applications. Noticeably, multiple kinds of research focus on evaluating the social characteristics based on the speech signal, spectral analysis of the recorded speech, and considering synthesis speech, which is seen as different from a spontaneous speech that occurs when people speak from an open mind.

Although the speaker's social characteristics in natural speech have not received much attention in a former investigation, this kind of study has not been proposed very often. Therefore, this study tries to deal with how listeners could perceive the speaker's social characteristics with a spontaneous conversation based on his/her voices.

2.1.1 A Taxonomy of Speaker Characteristics

Speech carries different types of information that worth researching. The essential type is words, which the speaker delivers to listeners. Besides, there is much other information that could be extracted and characterized from speech, like the speaker's emotions, gender, and the speaker's identity. This Task is known as speech recognition. The purpose of speaker recognition is to recognize the information about the speaker identity, in general, divided into two different classes :

1. Speaker verification: speaker verification is more concerned about verifying the identity of the speaker from his voices. Furthermore, control access to services such as telephone by banking and voice mail.

2. Speaker identification interested in automatically determining an unknown speaker's identity, mostly, it can be categorized into a category, open-set identification, and closed-set identification.

The taxonomy of speaker characteristics [21] and speaker recognition [13] are represented in fig 1. 2 and fig 2.2, respectively.
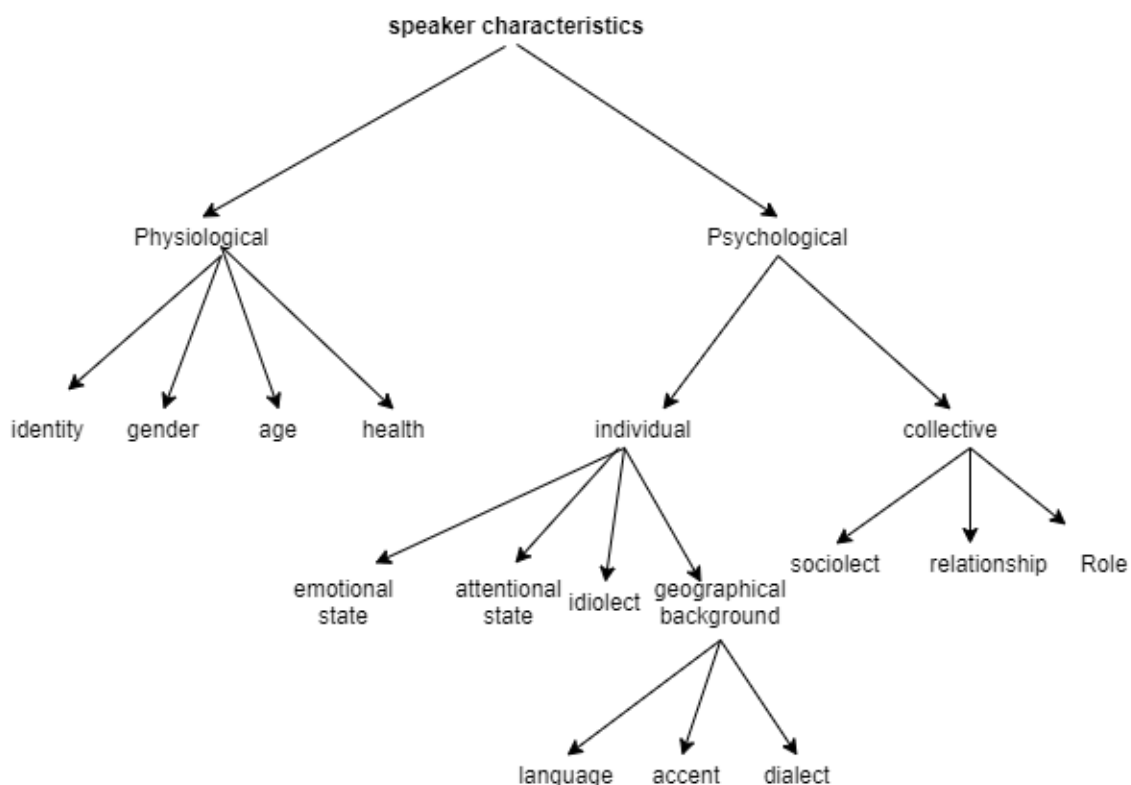
Figure 2.1: Taxonomy of speaker characteristics
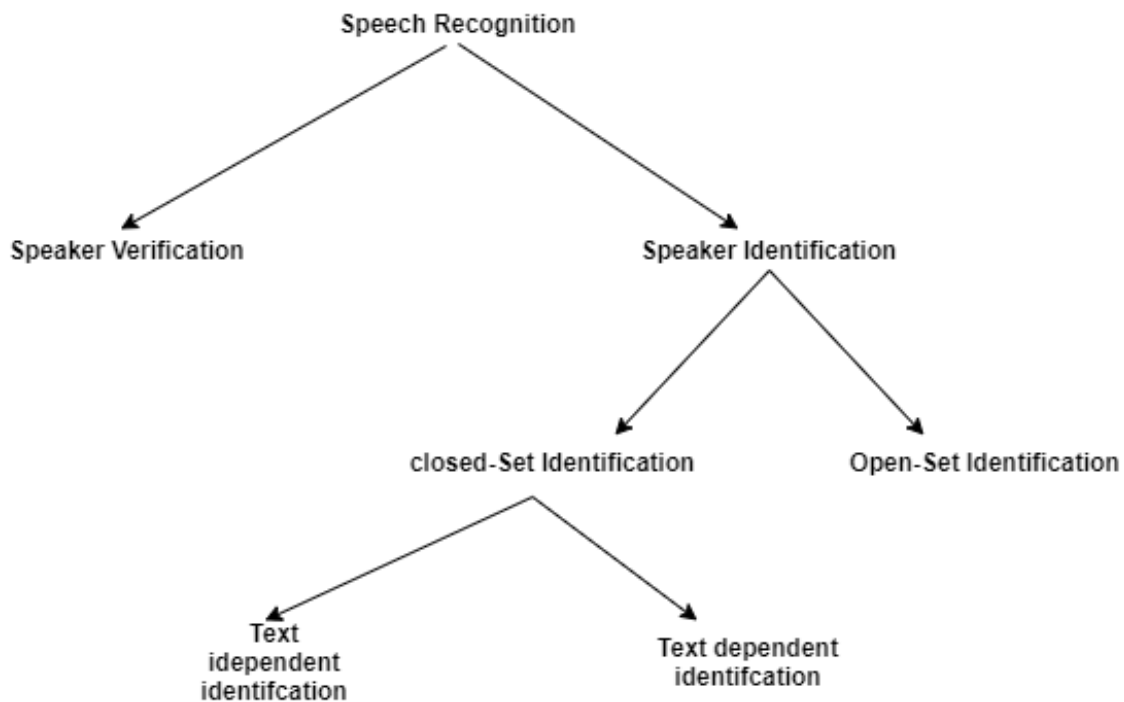
Figure 2.2: —Speaker Recognition

2.1.2   The Human Speech Production Mechanism

To study and analyze the speech, we first need to describe speech production and understand how one produces speech.

Speech production is the mechanism of uttering articulated sounds, i.e., how humans meaningful speech. The production of speech contains three levels of steps conceptualization, formulation, and articulation.

1. Conceptualization: In this process, creating speech attaches a desired concept to the spoken words to be output.

2. Formulation : The formulation stage contains the grammatical, lexical, and phonological encoding.

3. Articulation : In this step, the achievement of the articulatory is occurred to produce the speech.

Guenther & Frank (2006) described that "Broadly speaking, there are three main types of information involved in the production of speech sounds: auditory, somatosensory, and motor, which is represented in the temporal, parietal, and frontal lobes of the cerebral cortex, respectively."[11].

2.1.3 Speech Features

People's speech is full of properties and features that can be used to indicate a person's characteristics. Ambady and Robert (1992) stated that people moving, talking, facial expressions, posture, and speech all contribute to impressions about them[1].

Voice pitch and acoustic signals are the most features that have been addressed with personality judgments in multiple studies, and a study has been proposed by Borkowska & Pawlowski (2011) showed that men and women with lower voice pitch are assessed more dominant [3], whereas people with higher-pitched voices are judged as more nervous (Apple et al., 1979) [2].

2.1.4 Speech analysis

There are loft features that can be measured and give meaningful information in human speech. The extraction of these features from a speech signal is known as speech analysis. By interpreting time domain and frequency domain features like : Acoustic, pitch , and Voice quality are analyzed.

2.1.5 Applications of speech analysis

There are several approaches for applications speech processing which deal with speech, and his features, some of them are introduced in the following section:

1. Voice activity detection: these procedures pick out segments in a speech waveform.

2. Speech enhancement: this application domain enhances the speech signal's quality by removing the noise from speech.

3. Speech recognition: this work converts the speech signal to text.

4. Text to speech synthesis (TTS): this approach obtains natural speech from text and makes the sound of speech as natural as possible.

5. Speaker diarization and speaker recognition: in this context, the speech signal is broken into small segments belong to diverse speakers.

6. Audio source separation: this framework solves the problem of overlapping with speech and noise.

7. Speech modification: this method modifies speech features like emotion and tone.

8. Emotional speech classification: this framework evaluates emoting of the speech like happiness, sadness, and anxiety.

9. Keyword spotting: this technique indicts a specific keyword in the speech.

The analysis process of speech is based on three features include:

1. Duration: shows the histogram diagram of speech segments.

2. Time-domain: in the time domain the signal is represented, we can extract features like energy and zero-crossing rate.

3. Frequency domain : from the Frequency domain, we can obtain the most useful features of speech analysis.

2.1.6 Natural Speech vs Synthetic Speech

Spoken speech could be divided into two types :

1. Spontaneous speech: In this type of speech, people speak spontaneously, allowed to freely choose their own words, and also they have no prior idea about the topic which they are going to talk about.

   This kind of speech available in an interview, conversational telephone Speech, and dialog between people.

   More research has been conducted to study spontaneous speech. However, several challenges faced the researchers on how to figure out an emotion in spontaneous speech, which has motivated the researchers to do more studies in this field. Maekawa et al.(2015) claimed that "Many phoneticians and speech engineers are aware of the fact that 'real' human speech communication is quite

different from what they are analyzing in the laboratories, i.e., read-speech
"[15]. A work by Hoffman et al.(2010) investigated the temporal parameters
of spontaneous speech in three stages of Alzheimer's disease[12].

2. Synthetic speech
   A computer is used to generate synthetic speech, known as text to speech
   (TTS). More text-to-speech systems are designed to customize audio features.
   Manly, both synthetic speech and spontaneous speech are used for speech
   recognition systems and speech synthesis solutions.

2.2 Related Approaches

During the last years, the perceiving of speakers' social characteristics has drawn
both academia and the industry's attention to improve the customer services do-
main. A more recent study in this research is done by Gallardo & Weiss (2017).
They have investigated the effects of individuals' personality and of communica-
tion channels on speech likability in a speech utilizing a round-robin, and also they
have tried to measure and represent subjective person attributes measured from ob-
servers' first impressions based on speech only [10].
Gallardo and Sanchez (2016) have proposed a study to evaluate speakers' social per-
sonality, and they stated that this kind of study in demand in multiple systems[9].

A famous study (Stern et al.,2019) has been conducted to answer the question:
Does the voice carry vital information about a speaker's Personality? and tried
to address the significant relationship between people's personality traits and their
speech. The study has reported that ",the relationship between the personality traits
of the five-factor model of Personality, sociosexuality, and dominance is related to the
fundamental frequency (voice pitch) and formant frequencies p.2" [22]. Depending
on status (Brambilla et al.,2010) have perceived some personality traits like warmth
and competence.

Although there is an abundance of studies in this research area, there is still a
significant research gap in evaluating speakers' characteristics for different applica-
tion domains. This study tries to investigate the speaker's characteristics based on
the listeners' perceptions[4].

Chapter 3

Methodology and Methods

3.1   Methodology

As mentioned in the literature overview, predicting the speaker's social characteristics has been approached by many studies. Researchers have become increasingly engaged in the specific effects of social characteristics in our life.

This section will introduce the methods and procedures, and tools used to design the experiment, moreover, we will present the Conduct subjective evaluation and process analysis.

3.1.1   Research Questions and Hypothesis Testing

The present experiment was designed to test the hypothesis that the speaker's characteristics could be perceived from spontaneous speech based on listerine's rattlings. This research question could be express into two types of statistical hypotheses, the null hypothesis, which will be tested versus the alternative hypothesis.

(i) The null hypothesis states that: there is a significant correlation between the speaker's characteristics and spontaneous speech.

(ii) The alternative hypothesis states that: there is no significant correlation be-

tween speaker characteristics and spontaneous speech.

While we cannot measure the latent variables (factor is an observed variable) directly, we can measure them indirectly by using observed variables and doing a factor analysis to reducing the number of variables. In this experiment, to measure the latent variable, we have to measure some relevant observed variables in participants' responses with a 5-point scale.

3.1.2 Evaluation Dataset Model

Data collection is a clue aspect in building any speech analysis system, allows us to study and analyze its features. Moreover, developers can create more personalized and robust systems depending on this data.
Here we will highlight some of the procedures and research methodology that we have followed to collect the spontaneous speech samples, which explained and answered two questions, how the data was collected and how the data was processed.

The speech dataset was collected from native speakers, which addresses our needs, meets the research aims and objectives "Analysis of social speaker characteristics in natural speech".Furthermore, provides a valid result, such data could be found in human-to-human conversations, such as television interviews, talk show, and spontaneous telephone conversations, this kind of speech known as spontaneous conversation, in which speakers have used their own words, in other words, spoken are produced from the mind.

Collecting data is helpful for any study. The research questions predetermine the required data type for testing hypotheses. A 15 sample speech (10 female, 5 male) with two different scenario speech and conversation is utilized in this work. Each scenario consists of 5 segments of sentences, all between 5-7s duration. Besides secondary data, a primary dataset that directly addresses the research questions was collected from participants' responses.

3.1.3 Primary data processing

After collecting the raw data, we did a particular operation to obtain the speech samples and satisfy the statistical requirements.

Many tools can be used to process speech datasets. We have used a smart tool called audacity, which is open source and cross-platform, to extract our target data, in this case, is a guest's speech. We have done some treatments by using various tools, for instance, removing the laughs, overlapping the speech, and anchor's speech.

To accomplish the previous treatments in audacity, we have implemented some commands such as,

(i) Selecting the audio segmentation.

(ii) Apply some editing operations like cutting, merging, and removing the silent segments and overlapping between speakers.

(iii) Rearrange audio sergeants.

(iv) Export the results

(v) Save the audio with .mp3 format.

As a consequence, we have collected a natural speech dataset for native English speakers. It is approximately a one-hour audio duration with two distinct scenarios, "conversations and sentences," varying in length from 5 to 15 seconds.

3.1.4 Subjective Evaluation

In this experiment, self-reports methodologies are often used to collect participants' responses in observational studies and experiments. 2-groups of 39 participants are recruited(18 men, mean age $\pm$ SD = 25.7 $\pm$3.40, 21 women, 25.7 $\pm$ 3.60), based on ratings of 52 attributes, table 3.1 showed the most frequent attributes selected by the participants with on a 5-point scale, and table 3.2 gave some descriptive information about the participants.

A scatter plot in fig 3.1 showed an overview of female's conservation and sentences speech, and fig 3.2 showed a scatter plot of male's conservation and sentences speech.

Participants were given 45 min, with an average time of 50 sec for each question. The subjective assessment was collected from those participants with personal information, e.g. (age, gender) the average ratings calculated over subjective data can be found in this link: https://github.com/motasim-19/tu-proj/blob/master/Average ratings.xlsx. As the dataset was collected on an absolute scale, the ratings are normalized.

The observed variables are standardized(each variable's value is modified, so it has a mean of zero and a standard deviation of one),we used the z-score normalization.Then a significance analysis test was done on the collected data. The appropriate analysis was chosen based on the experiment's aim, statistical test, and factor analysis using statistical software package for the social sciences (IMB SPSS).
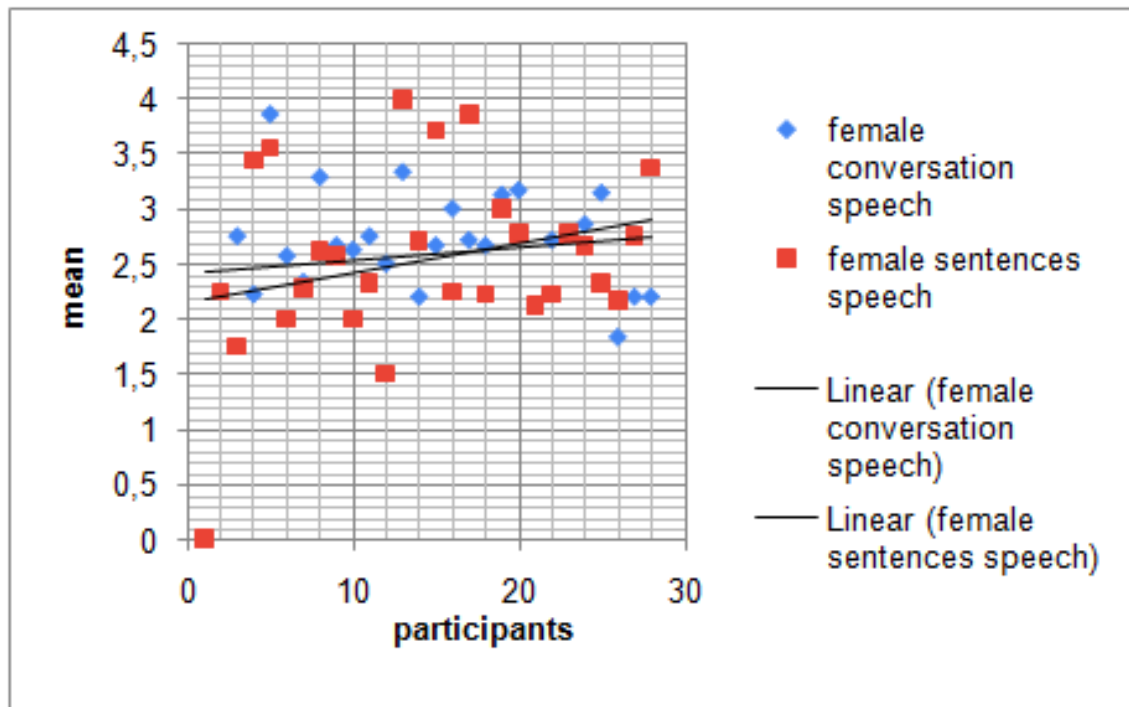
Figure 3.1: Scatter plot for female speech
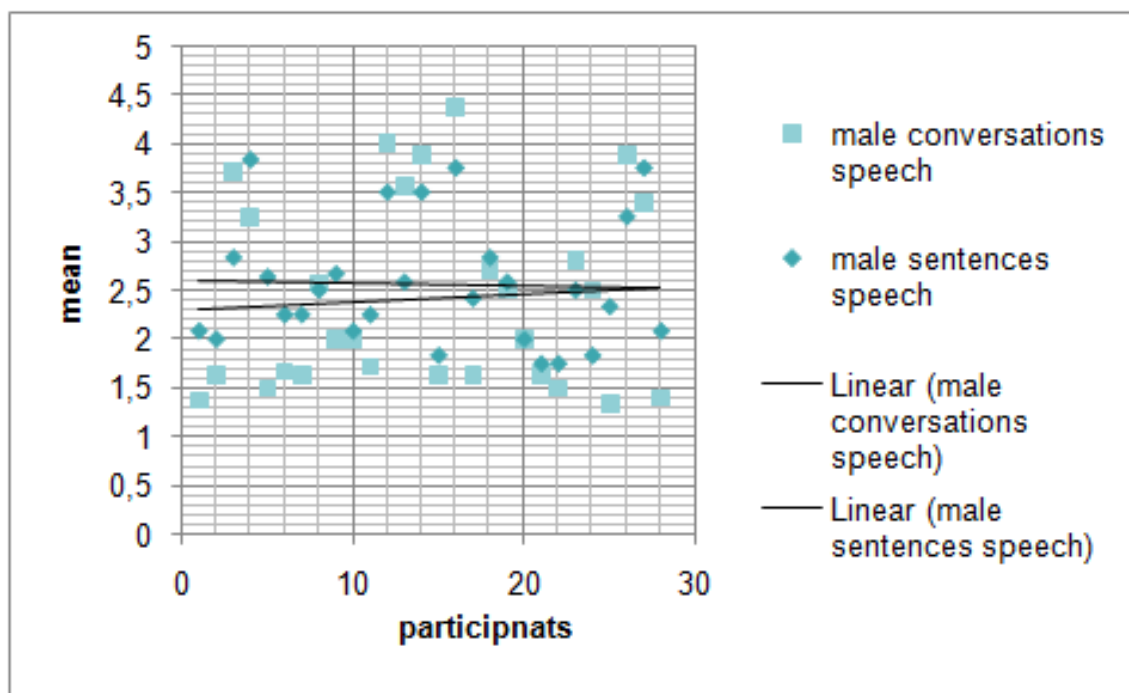
Figure 3.2: Scatter plot for male speech

Table 3.1: the most used attributes

subjective Attributes for ratting	
no	Attribute name
1	Responsibility
2	Friendliness
3	Introvert
4	Old
5	pleasant
6	Efficiently
7	Attractive
8	Proactively
9	Empathetic
10	Approachable
11	Appreciative
12	Arrogant
13	Emotionally
14	Worriedly
15	Reliable
16	Unstable
17	Intelligent
18	Enthusiastic
19	Energetic
20	Kindly
21	Calmly
22	Confident
23	Talkative
24	Agreeable
25	Thorough
26	Bored
27	Self-pitying
28	Forgiving

Table 3.2: Gender * type of speech Cross-tabulation

		type of Speech		
		Conversation	Sentence	Total
Gender	Female	12	9	21
	Male	9	9	18
Total		21	18	39

3.2 Data Analysis

3.2.1 Data Consistency and Reliability

Before extracting the factors, we have to verify the dataset, if it is suitable for factor analysis or not and have similar responses. In other words, among the observed variables, which variables consist of a set on underlying variables " factors". Two tests were conducted. The first test is Kaiser-Meyer-Olkin (KMO) and Barliest test. The second test is the questionnaire's reliability to check if the questioners in the survey measure the same underlying construct that by cronbach alpha test.

3.2.2 KMO and Barliest test

Henry Kaiser (1970) proposed a proposal to measure sampling adequacy (MSA) of factor analytic data matrices [18] ,kaiser test was done to measure the sampling adequacy. The optimal Kaiser-Meyer-Olkin (KMO) values are between .8 and 1.

As shown in the table 3.3, the KMO value is 0.6, indicating that the sampling is adequate. Barliets test was done to check if the dataset is significant or not. In particular, it tests the hypothesis that the correlation matrix is an identity matrix. According to the p-value, if the p-value <0.05, then the null hypothesis will be rejected, otherwise is accepted. As noticed from table 3.3 the P-value =0.0001 which is <0.05 "the most Significance Level (Alpha=0.05) used in statistical test", Therefore, it is conduct that the variables are useful, so we can make useful factors, which means at least two questions that have positively correlated.

Table 3.3: Measures for Assessing the Correlation Matrix

Measure	Value
Bartlett's Test of Sphericity	P <.0001
Kaiser-Meyer-Olkin Test of Sampling Adequacy	.591

3.2.3   Cronbach's Coefficient Alpha

Cronbach's alpha is the most statistical test (Cronbach, 1951) have been designed test the internal consistency reliability among items in scales[6].
Keith & Taber (2017) mentioned that "reliability coefficient value is alpha values were described as following excellent (0.93–0.94), strong (0.91–0.93), reliable (0.84–0.90), robust. (0.81), fairly high (0.76–0.95), high (0.73–0.95), good (0.71–0.91), relatively high (0.70– 0.77), slightly low (0.68), reasonable (0.67–0.87), adequate (0.64–0.85), moderate (0.61– 0.65), satisfactory" [23].Table 3.4 summarizes the Cronbach's alpha coefficients values, and we observed the value of Cronbach's coefficient >0.7, which is considered as appropriate for doing factor analysis.

Table 3.5 represents the chromebox apha if the item is deleted from the scale. When we are going through this table , if items (8), (10), (12), (14), and (26) are deleted, Cronbach's coefficient of the corresponding factor increases slightly.

Table 3.4: Reliability Statistics test

Measure	Value
Cronbach's Alpha	.923
Cronbach's Alpha Based on Standardized Items	.928
N of Items	23

Table 3.5: Internal consistency of items test

items	Cronbach's Alpha if Item Deleted
Responsibility	.0.914
Friendliness	0.912
Introvert	0.914
Old	.0.914
pleasant	.0.913
Efficiently	.0.914
Attractive	.0.911
Proactively	.0.915
Empathetic	0.912
Approachable	0.911
Appreciative	.0.916
Arrogant	.0.926
Emotionally	0.913
Worriedly	.0.915
Reliable	.912
Unstable	0.911
Intelligent	0.911
Enthusiastic	0.914
Energetic	0.913
Kindly	.0.912
Calmly	.0.913
Confident	.914
Talkative	.914
Agreeable	.913
Thorough	.915
Bored	.0.917
Self-pitying	.914
Forgiving	.914

3.3 Conducting Factor Analysis

3.3.1 Factor Analysis Basic Idea

The factor analysis's basic idea is to figure out a set of the underlying variable from a collection of observed variables. Hence in this study, we try to break the interrelationships between items into multiple components, so factor analysis is appropriate for this purpose of analysis. Figure 3.3 demonstrates a set of variables underlying into specific factor[16].
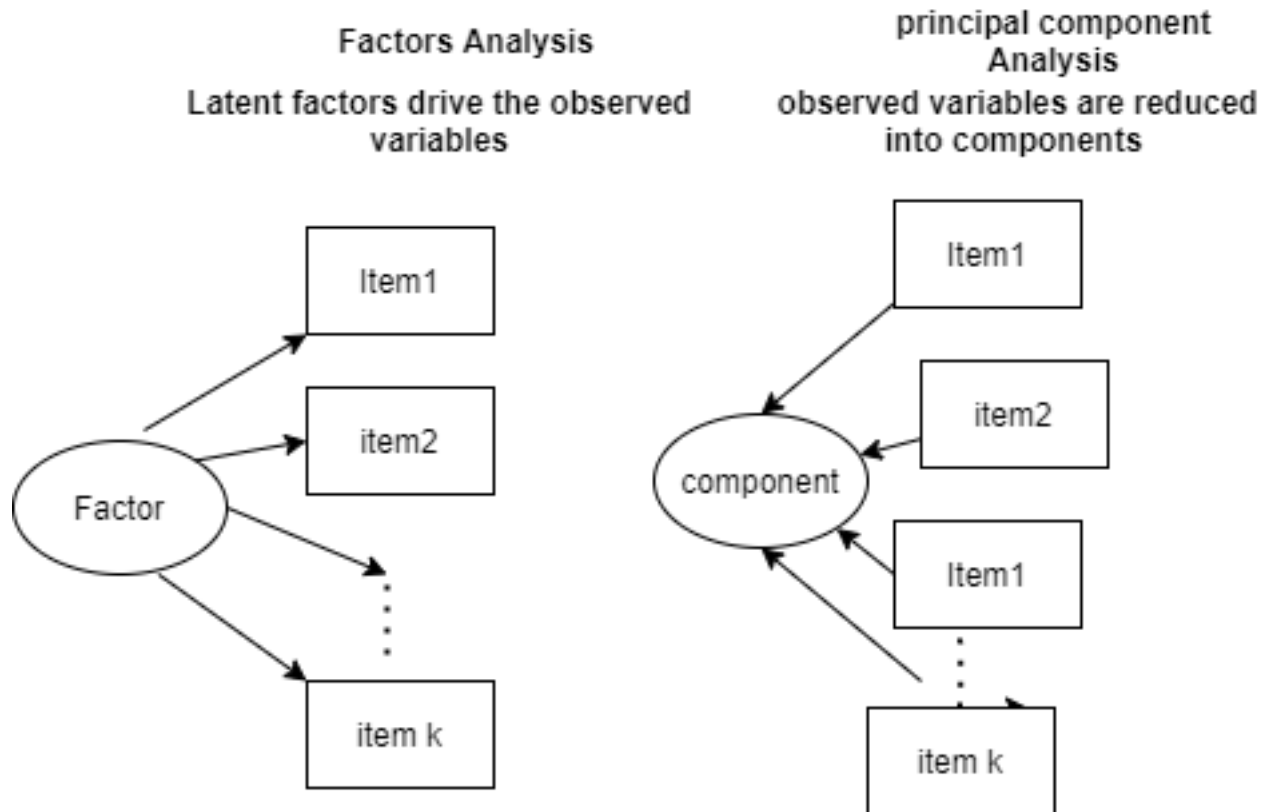
Figure 3.3: Factor analysis

Factor analysis could be classified as an exploratory and confirmatory factor. Exploratory factor analysis (EFA) is preferred in this study for two reasons:

1. The first reason is that we do not know the number and nature of the underlying factors (that latent variable).

2. A second reason, as we want to identify the factor structure underlying a set of data, is suitable for exploratory factor analysis.

Consequently, by carrying out the exploratory factor analysis (EFA) on observed variables, we can determine the number of constructs measured by the questionnaire, furthermore the nature of those constructs. Factor analysis was conducted in a sequence of steps, factor extraction, and factor rotation.In the following link a results of factors analysis with different rotation methods could be presented https://github.com/motasim-19/tu-proj/blob/master.

3.3.2   Retain The Number of Factor

There are some ways to determine the number of components to retain that have been proposed. In this experiment, the most widely used extraction method, "fixed number of factor and eigenvalues," is used for determining the number of factors with suppressed threshold values of .5; however, some studies state that eigenvalues are a traditional method for extracting factor compare to others approach, like parallel analysis which based on random data generation factors. After retaining the initial number of factors, but how many meaningful factors should be extracted?, according to kaiser-guttman criterion (kaiser 1960) any factors with an eigenvalue greater than one is retained [24].

Although Çokluk & Koçak (2016) proposed a new idea to extract the number of a factor called parallel analysis based on random data generation [5].

3.3.3   Factor Extraction

The big problem faces by designing a survey is a redundancy of variables. Therefore, the number of variables should be reduced to a set of observed variables to deal with the problem we used with principal components as the extract method.

3.3.4   Factor Rotation

We want to show the correlation between the components and use this information to interpret the components. Factor rotation was done to obtain a simple structure which interrupt how the items load on each factor. Replicate of rotation is done and the number of factors was determined by 3,4,5 with different types of rotation methods such as varimax(an orthogonal solution that results in uncorrelated),equimax and promax rotations the results of factor analysis on subjective evaluations can be found here https://github.com/motasim-19/tu-proj/blob/master.
The results displayed that varimax rotation with three fixed number of factors gave simple structure and easy to interrupt, that most the variables have high loading on one component and near-zero on other components, the items were retained when the main loading was greater than 5 and the difference between the main loading and the complex loading was at least 0.02. Using the varimax rotation method

with combining with other rotation methods a full data and meaningful, significant factors loading was obtained and had fewer cross loading. Each variably loads high onto only one factor.

Table 3.6: Factor loading for Sentences speech(male group) **(*) indicates that items belongs to both female and male speakers**

items	factor1	factor2	factor3
Kindly ∗	.86		
Forgiving ∗	.84		
Talkative ∗	.83		
Empathetic	.82		
Approachable ∗	.80		
Thorough	.59		
Agreeable	.51		
Energetic		.91	
Calmly∗		.89	
Reliable		.88	
Attractive		.87	
Emotionally		.80	
pleasant∗		.78	
Unstable		.60	
Old∗		.54	
Enthusiastic			.75
Intelligent			.74
Confident			.72
Efficiently			.63

Table 3.7: Factor loading for conversation speech (Female group)

items	factor1	factor2	factor3
Self-pitying	.80		
Enthusiastic	.78		
Talkative	.78		
Introvert	.77		
Approachable	.74		
Forgiving	.74		
Emotionally	.67		
Efficiently	.54		
Kindly	.54		
Intelligent		.94	
Calmly		.92	
Old		.85	
Confident		.76	
pleasant		.75	
Agreeable			94
Thorough			78
Empathetic			71
Unstable			71
Responsiblty			56

3.4   Independent Statistic t-test

Independent t-test was conducted at a certain significance level of α=0.5 to reject the null hypothesis H_0 or accept it, if there is a statistically significant difference between female and male groups in both conversations and sentence scenarios, there is no significant difference between them.

As seen in table 3.9 the result of the independent t-test indicates the levens's test for the quality of variances is significant (i.e., p-value 0.26 >0.05). Therefore we consider the "Equal Variances assumed values". Then we look at sig(2-tailed) value, 0.327 >0.05, that indicates that the null hypothesis is accepted, and we inferred there is no significant difference between female group and male group (mean =85, SD= 0.13 and group male (mean 0.90, SD=0.06) when the participants perceived conversations speech.

Table 3.8: Statistical info for rating conversations group

Gender	N	Mean	Std. Deviation	Std. Error Mean
female	9	0,85	0,138	0,04606
male	9	0,90	0,0677	0,02258

Table 3.9: t-test for Equality of Means

		F	Sig	t	df	Sig.(2-tailed)	Mean Difference
MeanRating	Equal variances assumed	1.334	0.265	-1.011	16	0.327	-0.05189
MeanRating	Equal variances not assumed			-1.011	11.637	0.332	-0.05189

A synonym test t-test was conducted to test, whereas there is a significant difference between female and male speech sentences. As seen in table 3.10 and 3.11,the result

presented that levens's test for quality of variances is significant(i.e., p-value 0.03 <0.05), therefore we consider the the "Equal Variances not assumed values", since the sig(2-tailed) 0.02 <0.05 ,that denoted can the null hypothesis is rejected,so there is a significant difference between female group and male group (mean =85,SD= 0.086 and group male (mean 0.94,SD=0.036) speech is perceived.

Table 3.10: Statistical info for rating sentences group

Gender	N	Mean	Std. Deviation	Std. Error Mean
female	12	0,9419	0,0364	0,01053
male	9	0,8576	0,086	0,02868

Table 3.11: t-test for Equality of Means

		F	Sig	t	df	Sig.(2-tailed)	Mean Difference
MeanRating	Equal variances assumed	5,318	0,033	3,065	19	0,006	0,08426
MeanRating	Equal variances not assumed			2,758	10,170	'0,020	0,08426

Chapter 4

Results and Discussion

4.1 Background

After the factor analysis was conduct to investigate the latent variable, we have to interpret the high loading factor. Steven (200) provides some guidelines in this matter, that is, we count only a loading if the absolute value $>.5$ and the difference between the main loading and the complex loading is at least 0.02 [22].

4.2 Interrupting The Factor Loading

From the male speaker's analysis result, it could be observed that component1, component2, and component3 account for 67.61% (37.9%,16.78%,12.87%, respectively) of variance among the observed variables, see table 4.1. It is also clear that from the second female speaker scenario component1, component2,and component3,account for 73.02% of the total variance (50.77%, 11.98% ,10.27 % respectively), see table 4.2.

4.2.1 Perceptual Dimensions Extraction

Refer to table 3.6 and table 3.7 in chapter 3 which interrupted the factor loading for both Scenario female and male speech, respectively. We noticed that seven and nine items are underlying in the first component in male and female speech, respectively. Among them, four attributes are found in both female and male groups.

We also see eight and five items underlying the second component in male and female speech. Among them, three attributes are found in both female and male groups.

There are four other and five items observed in the third component, respectively. One item was found in both male and female speech. The first component will be correlated with some of the observed variables. The second component will be correlated to some observed variables that did not display in strong component1 and the third component. (O'Rourke & Hatcher) [17].

By looking at the nature of variables that are loaded under the same factor, the nature of that factor can be recognized. It can be identified from table 3.6 and table 3.7 a common theme that loads high significantly onto factor1 are seemed related to one Social Characteristics, therefore, it reasonable to label the items are loaded on factor1 warmth, the items that load high significantly onto factor2 seem to be measured the same construct social characteristic, therefore it possible to label them as competence.

Finally, it could be inferred that the items that load onto factor3 contain social characteristics attributes.

Table 4.1: Initial Eigenvalues for male

Component	%of Variance	Cumulative %
1	37.95	37.95
2	16.78	54.74
3	12.87	67.61
4	7.26	74.8
5	6.4	81.2
6	4.9	86.28
7	4.40	90.69
8	2.86	93.5

Table 4.2: Initial Eigenvalues for female

Component	%of Variance	Cumulative %
1	50.7	50.75
2	11.98	62.75
3	10.27	73.02
4	6.50	79.53
5	5.62	85.16
6	4.58	89.74
7	3.950	93.69
8	2.07	95.76

Based on the t-test results in table 3.9 and table 3.11, we can stated that :

1. Female speakers perceived the same as and male speaker in the conversations speech.

2. Female speakers perceived differently from the male speaker in the sentences speech. In other words, females have predicted a statistically significantly higher mean than males.

Chapter 5

Conclusion

In sum, the study has demonstrated that social speaker characteristics could be perceived from spontaneous speech based on the listener's responses, and this leads to good results,there is a significant correlation between speaker characteristics and his voices. A findings showed two core dimensional of social perception can be recognized ,when analyzing the social characteristics of speaker namely, "warmth and competence".

Consequently, the factor analysis resulted that there are four attributes represent warmth expressly, "Kindly, "Forgiving, "Talkative," "Empathetic, "Approachable" and three attributes expressed competence, namely, "Calmly," "pleasant," "Old" in both female and male group.Verifying that these set of attributes tend to available in many human characteristics.

Prominently, there was a statistically significant difference between female and male groups when evaluating sentence speech(P-values 0.02).
In reverse, there was no significant difference between the female and male groups when evaluating conversation speech(P-values 0.327).The study faced common changelings and limitations issues in practice such as, it is tricky to estimate in advance which attributes are appropriate to serve in recognition of the latent variable. Also, it is difficult to identify the most useful approach to retain the number of factors. Furthermore, it is hard to collect a large sample size due to insufficient participants who can make reasonably professional responses.

Finally, to address and characterize more advanced speaker characteristics, we

need to consider some other speaker traits like speech rate, accent, and the quality of the speech, which have a significant impact and influence in speech perception estimating.

34

Chapter 6

Appendix A : questionnaire

of the questions and statements included during the in questionnaire :

Q1: Would you like to rate the speech sample on a scale for the attribute

Statements: Continuous scale for the assessment of the attribute attribute.

Sampe of survey

https://docs.google.com/forms/d/e/1FAIpQLScBJcFReYf0d-JJNTwEwIcvW3afV9j08sg7bqQn0lErVWJ3xA/viewform?usp=sf$_link$.

Chapter 7

Appendix B :Supplementary files

participant's response raw data

https://docs.google.com/spreadsheets/d/1APz6CuXiJmF3DNO8Bj-PpoE8ciaWNBxEalYYmP5FG1g/edit?usp=sharing

Bibliography

1. Ambady, N. & Rosenthal, R. Thin slices of expressive behavior as predictors of interpersonal consequences: A meta-analysis. *Psychological bulletin* **111,** 256 (1992).

2. Apple, W., Streeter, L. A. & Krauss, R. M. Effects of pitch and speech rate on personal attributions. *Journal of personality and social psychology* **37,** 715 (1979).

3. Borkowska, B. & Pawlowski, B. Female voice frequency in the context of dominance and attractiveness perception. *Animal Behaviour* **82,** 55–59 (2011).

4. Brambilla, M., Sacchi, S., Castellini, F. & Riva, P. The effects of status on perceived warmth and competence. *Social Psychology* (2010).

5. Çokluk, Ö. & Koçak, D. Using Horn's Parallel Analysis Method in Exploratory Factor Analysis for Determining the Number of Factors. *Educational Sciences: Theory and Practice* **16,** 537–551 (2016).

6. Cronbach, L. J. Coefficient alpha and the internal structure of tests. *psychometrika* **16,** 297–334 (1951).

7. Diener, E. & Lucas, R. E. Personality traits. *General Psychology: Required Reading,* 278 (2019).

8. Flanagan, J. L. *Speech analysis synthesis and perception* (Springer Science & Business Media, 2013).

9. Gallardo, L. F. & Weiss, B. *Speech Likability and Personality-Based Social Relations: A Round-Robin Analysis over Communication Channels.* in *Interspeech* (2016), 903–907.

10. Gallardo, L. F. & Weiss, B. *Towards Speaker Characterization: Identifying and Predicting Dimensions of Person Attribution.* in *INTERSPEECH* (2017), 904–908.

11. Guenther, F. H. Cortical interactions underlying the production of speech sounds. *Journal of communication disorders* **39,** 350–365 (2006).

12. Hoffmann, I. *et al.* Temporal parameters of spontaneous speech in Alzheimer's disease. *International journal of speech-language pathology* **12,** 29–34 (2010).

13. Hossain, M., Ahmed, B. & Asrafi, M. *A real time speaker identification using artificial neural network* in (Jan. 2008), 1–5. ISBN: 978-1-4244-1550-2.

14. Little, A. C., Jones, B. C. & DeBruine, L. M. Facial attractiveness: evolutionary based research. *Philosophical Transactions of the Royal Society B: Biological Sciences* **366,** 1638–1659 (2011).

15. Maekawa, K., Koiso, H., Furui, S. & Isahara, H. *Spontaneous Speech Corpus of Japanese.* in *LREC* (2000), 947–9520.

16. Matsunaga, M. How to Factor-Analyze Your Data Right: Do's, Don'ts, and How-To's. *International journal of psychological research* **3,** 97–110 (2010).

17. O'Rourke, N. & Hatcher, L. *A step-by-step approach to using SAS for factor analysis and structural equation modeling* (Sas Institute, 2013).

18. Revelle, W. KMO: Find the Kaiser, Meyer, Olkin Measure of Sampling Adequacy.

19. Schötz, S. *Perception, analysis and synthesis of speaker age* (Lund University, 2006).

20. Schuller, B. *et al.* A survey on perceived speaker traits: Personality, likability, pathology, and the first challenge. *Computer speech & language* **29,** 100–131 (2015).

21. Schultz, T. in *Speaker classification I* 47–74 (Springer, 2007).

22. Stern, J. *et al.* Do voices carry valid information about a speaker's personality? (2019).

23. Taber, K. S. The use of Cronbach's alpha when developing and reporting research instruments in science education. *Research in Science Education* **48,** 1273–1296 (2018).

24. Velicer, W. F. Determining the number of components from the matrix of partial correlations. *Psychometrika* **41,** 321–327 (1976).