

Vocal quality and personality stereotypes

Jukić, Alan

Undergraduate thesis / Završni rad

2022

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **Josip Juraj Strossmayer University of Osijek, Faculty of Humanities and Social Sciences / Sveučilište Josipa Jurja Strossmayera u Osijeku, Filozofski fakultet**

Permanent link / Trajna poveznica: <https://um.nsk.hr/um:nbn:hr:142:493057>

Rights / Prava: [In copyright](#) / [Zaštićeno autorskim pravom.](#)

Download date / Datum preuzimanja: **2025-03-12**



FILOZOFSKI FAKULTET
SVEUČILIŠTE JOSIPA JURJA STROSSMAYERA U OSIJEKU

Repository / Repozitorij:

[FFOS-repository - Repository of the Faculty of Humanities and Social Sciences Osijek](#)



Sveučilište Josipa Jurja Strossmayera u Osijeku
Filozofski fakultet Osijek
Dvopredmetni sveučilišni preddiplomski studij engleskog jezika i književnosti i
njemačkog jezika i književnosti

Alan Jukić

Kvaliteta glasa i stereotipi osobnosti

Završni rad

Mentorica: doc. dr. sc. Alma Vančura

Osijek, 2022.

Sveučilište Josipa Jurja Strossmayera u Osijeku
Filozofski fakultet Osijek
Engleski jezik i književnost
Dvopredmetni sveučilišni preddiplomski studij engleskog jezika i književnosti i
njemačkog jezika i književnosti

Alan Jukić

Kvaliteta glasa i stereotipi osobnosti

Završni rad

Znanstveno područje: humanističke znanosti

Znanstveno polje: filologija

Znanstvena grana: anglistika

Mentorica: doc. dr. sc. Alma Vančura

Osijek, 2022.

J.J. Strossmayer University of Osijek
Faculty of Humanities and Social Sciences
Double Major BA Study Programme in English Language and German Language
and Literature

Alan Jukić

Vocal quality and personality stereotypes

Bachelor's Thesis

Supervisor: Alma Vančura, Ph.D. Assistant Professor

Osijek, 2022

Josip Juraj Strossmayer University in Osijek
Faculty of Humanities and Social Sciences
Department of English
Double major BA study Programme in English Language and Literature and
German Language and Literature

Alan Jukić

Vocal quality and personality stereotypes

Bachelor's thesis

Scientific area: humanities
Scientific field: philology
Scientific branch: English studies

Supervisor: Alma Vančura, Ph.D. Assistant Professor

Osijek, 2022

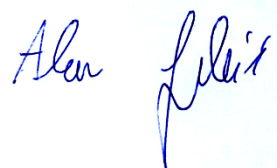
IZJAVA

Izjavljujem s punom materijalnom i moralnom odgovornošću da sam ovaj rad samostalno napravio te da u njemu nema kopiranih ili prepisanih dijelova teksta tuđih radova, a da nisu označeni kao citati s napisanim izvorom odakle su preneseni.

Svojim vlastoručnim potpisom potvrđujem da sam suglasan da Filozofski fakultet Osijek trajno pohrani i javno objavi ovaj moj rad u internetskoj bazi završnih i diplomskih radova knjižnice Filozofskog fakulteta Osijek, knjižnice Sveučilišta Josipa Jurja Strossmayera u Osijeku i Nacionalne i sveučilišne knjižnice u Zagrebu.

U Osijeku, 30. Kolovoza 2022.

Alan Jukić, 0122226884



Summary

Vocal quality or vocal trait is unique for each person and represents a natural trait of every person. It depends on many different aspects, such as age, sex, personality type, medical conditions and more. Moreover, it can be altered and manipulated by the speaker, creating an illusion of the vocal quality the listeners can hear. The perception of the vocal quality has not changed much since the past; however, there are interesting takes on vocal perception that differ from modern vocal qualities. People tend to analyze speakers' voices and, in the process, sometimes add stereotypes. This often happens subconsciously and is a common trait among people worldwide. Combine that with unique vocal features every person has and the result is the envisioning of personal characteristics through the sound of the voice. This paper will give an overview of different phonatory settings, history of registers, perception of voices in modern times and some medical conditions that influence our perception of voice.

Key words:

Vocal quality, personality, stereotypes, vocal characteristics, history

Table of contents

- 1. Introduction 1
- 2. Vocal quality 3
- 3. Phonatory settings 5
 - 3.1. Voiced and voiceless features 5
 - 3.1.1 Neutral mode 7
 - 3.1.2. Modal Voice 7
 - 3.1.3. Vocal fry 8
 - 3.1.4. Breathy voice 9
 - 3.1.5. Harsh voice 9
 - 3.1.6. Falsetto 10
 - 3.2. Vocal quality combinations 11
- 4. Vocal quality, paralinguistics and cultural differences 12
- 5. History of registers and vocal quality 14
 - 5.1. History of registers 14
 - 5.2. History of vocal quality 17
- 6. Personality stereotypes 18
- 7. Altered vocal quality 20
 - 7.1. Medical conditions 20
- 8. Vocal quality in modern times 22
- 9. Conclusion 25
- References 26

1. Introduction

Vocal quality is difficult to define, as there are many factors which influence its quality. How important vocal quality is can be seen by observing various aspects of everyday life, and it shows speaker's inner strength, emotions and feelings. The paralinguistics of speech are the non-linguistic effects, which indicate traits such as gender or personality and may consequently bias our choice "of whom to like, whom to trust, and with whom to do business" (Baird, et. al., 2018: p. 277). Paralinguistics helps us to understand what the difference between sarcasm and sincerity, honest concern and teasing, happiness and annoyance is. Although paralinguistic features are innate, there are many outside factors that influence and model them, and there is still a way for the speaker to control their voice in many different ways.

This paper will present the research done on the vocal quality and its influence on the listener's perception. It will also show the usage of different phonatory settings and its influence on the voice. The aim of this paper is to show how different phonatory settings create a various impressions on the listener and how an altered voice can manipulate the speaker's image.

People change their voice depending on the roles they have in life or in a daily routine. Many great speakers such as Hitler used their voice to make people listen to them and follow their beliefs. Here is an example of how Hitler used his voice to influence the public.

Introductory part of the speech would start with a deeper tone, slower pace and a quieter voice and would thus leave, at least apparently, an impression of a sane person. On the contrary, when he would paraphrase the opponents or call the Germans to unity, Hitler would become visually and audibly more engaged. He would raise his intonation very rapidly, in the span of a single sentence. He would then increase the pace of speaking, which would be followed by staccato speech or speech with a strong stress (giving importance to every single word) and strong laryngeal pulses, followed by guttural speech with posterior, drawn back articulation. That type of speech signals determination, insensitivity, while the volume increase complements the contents of the speech. His guttural voice would then become even more closed off and hoarse. He would finish his engaging part

decisively, with a strong rising, or more often, a strong falling intonation¹. (Legati, 2008: p. 210-211)

This paper is structured as follows. Section 2 focuses on the general discussion on vocal quality, its definition and connection to the paralanguage. Section 3 covers the definition of the phonatory settings and their individual features. Section 4 analyses the perception of the speaker and how it changes according to the geographical location of the speaker. Section 5 dwells into the historic analysis of the vocal quality and different registers. Section 6 analyzes the attribution of different personality stereotypes while listening to a speaker. Section 7 covers alterations in vocal quality, which can be intentional or stem from a medical condition. Section 8 covers vocal quality in modern times and technological advancements with a glare into the foreseeable future. Finally, section 9 gives the conclusion of the paper.

¹Translated from Croatian by Jukić

2. Vocal quality

A definition of a vocal quality is different depending on the author. Speakerhub (n.d.) says that the vocal qualities are certain speech variations which help to distinguish different voices. They relate to different factors such as speaker's breath, air pressure, phonation and other factors. On the other hand, Kreiman et al. said (2003: p.2) that "defining and quantifying vocal quality remains an elusive goal". However, Laver defines it as (1980: p. 1)

Voice quality is conceived here (...) in a broad sense, as the characteristic auditory colouring of an individual speaker's voice, and not in the more narrow sense of the quality deriving solely from laryngeal activity. (...) Perceptually, voice quality in this broad interpretation is a cumulative abstraction over a period of time of a speaker-characterizing quality, which is gathered from the momentary and spasmodic fluctuations of short-term articulations used by the speaker for linguistic and paralinguistic communication.

As a part of paralinguistic speech, vocal quality is present at all times while communicating to other people. Paralanguage is defined as

the nonverbal voice qualities, voice modifiers and independent utterances produced or conditioned in the areas covered by the supraglottal cavities (from the lips and the nares to the pharynx), the laryngeal cavity and the infraglottal cavities (lungs and esophagus), down to the abdominal muscles, as well as the intervening momentary silences, which we use consciously or unconsciously supporting or contradicting the verbal, kinesic, chemical, dermal and thermal or proxemic messages, either simultaneously or alternating with them, in both interaction and noninteraction. (Poyatos, 1993: p. 6)

It helps recognise the speaker as well as their affective states. "Being a parameter that is difficult to control intentionally, voice quality serves as a strong indicator of the affective states of the speaker [...], and is perhaps the most strongly recognised feature of paralinguistic speech, albeit subconsciously" (Esposito et al., 2007: p. 121). This implies that people can intuitively recognize whether the speaker is self-confident, fearful, angry, sad, etc. This can be very helpful when talking to people. It may be a difference maker between a person getting a job or not at a job interview. It

is important to note that the perception of the paralinguistic speech can be very subjective. Different people can interpret the same speaker differently. They understand the speaker through their own eyes and ears and may associate the speaker's paralinguistic features with something from their previous encounters.

Taken together, the sum of all the labels describe many aspects of the speech, but individually they can be difficult to compare. For example, labeller A may determine that the speaker is 'speaking softly', labeller B that she is 'being kind', labeller C that she is 'acting cute', and so on. (Esposito, et. al., 2007: p. 122)

These speaker characteristics which different labellers understand in different ways, may all be true or may all be false for the speaker. Those are merely labeller's perceptions. According to the authors (ibid.), this confusion may appear due to the listeners taking different points of view and analyzing the speaker either according to their appearance, or their pragmatic function (the information that is attempted to be transferred) or even the mechanics of their speaking style (manuscript style, memorized style, impromptu style and extemporaneous style).

3. Phonatory settings

Speaking about voice and vocal tendencies, we have to first mention the term “register”. Just as with vocal quality, it is not that easy to define this term in a single sentence. Laver (1980) had the same issue arguing that most of the so-called definitions are merely descriptions which do not pay adequate tribute to the term itself.

Even in the case of modal voice, it is only a description, and remains far from a definition: the issue is merely hedged by saying that modal voice is the type of vocal fold vibration which phonetic theory assumes takes place in ordinary voicing, when no specific feature is explicitly changed or added. It is very hard to construct a satisfactory, short definition of modal voice... (Laver, 1980: p. 95)

Laver (1980) also states that the register covers not only the laryngeal behaviour (area of the throat containing vocal cords) but also a whole constellation of activities at various levels of the vocal tract. This would imply that the register encompasses not only the vocabulary used in a spoken language (supralaryngeal activities) but also the types of phonation (phonatory activities). Laver also explains what phonatory settings are. “The domain of phonatory settings is limited by the same criterion that was applied to supralaryngeal settings: only those settings which can potentially be controlled by any speaker with a normal vocal apparatus will be admitted into the descriptive phonetic scheme (Laver, 1980: p. 93).

3.1. Voiced and voiceless features

The basic classification of the phonatory settings is divided into voiced and voiceless settings, and from then on, into many other categories. In the table below are shown those settings that will be presented in the paper.

Table 1. Voiceless and voiced phonation²

voiceless	nil	whisper			
voiced	modal	creak	breathy	harsh	falsetto

According to Ladefoged (1971), voiceless features are realized through vocal cords being apart at the posterior end between the arytenoid cartilages. Ladefoged (ibid.) also says that the glottis is so far apart that it cannot be set in vibration and therefore cannot create true sound. Voiced features are described by Ladefoged (1971: p. 7) as follows:

In the formation of a voiced sound the vocal cords are adjusted so that they are almost touching along their entire length. The result of air flowing through this constriction is a suction effect which draws the vocal cords together. But as soon as they are together there is no flow, and consequently no action pulling them together; so they come apart and release the pressure which has been built up beneath them. But when they are apart they are again subject to the suction caused by the outgoing air. So the cycle repeats itself, producing the regular vibrations known as voice.

Laver (1980: p. 121) says the following about whispery voice:

In weak whisper, the triangular opening can be fairly long, including part of the ligamental as well as the cartilaginous glottis. With increasing intensity, the glottis is increasingly constricted until only the cartilaginous section remains just open. Taken together, these factors suggest low adductive tension, and moderate to high medial compression

Whispery voice is mostly used for secretive talks or to avoid disturbance.

On the other hand, the voiced features produce sound due to the vocal folds vibrating.

²<https://www.ims.uni-stuttgart.de/institut/arbeitsgruppen/ehemalig/ep-dogil/EGG/page10.htm>

3.1.1 Neutral mode

In presenting all the phonatory settings, we will first start with the neutral mode. “The neutral mode of phonation is one where the vibration of the true vocal folds is periodic, efficient, and without audible friction” (Laver, 1980: p. 94). This means that the neutral mode is the one where the message can be transferred from a speaker to a listener in the most effective way, with no obstructions or problems. This does not mean that every listener has issues in understanding different modes of phonatory settings, but just means that the average listener has the greatest ability to understand the neutral mode of spoken language.

3.1.2. Modal Voice

The main phonation is the afore mentioned neutral or also called modal voiced phonation. Laver (1980: p. 95) describes modal voice as “...the type of vocal fold vibration which phonetic theory assumes takes place in ordinary voicing, when no specific feature is explicitly changed or added. “Laver (ibid.) says that “the production of modal voice is thus carried out with only moderate adductive tension and moderate medial compression, with moderate longitudinal tension when the fundamental frequency is in the lower part of the range used in ordinary conversation“. Laver (1980) writes about the invention of the term modal phonation, saying the first person to use it was Hollien in 1971. “...the term adopted here for the neutral mode of vibration, modal voice is Hollien's (...) he says he chose the term 'modal' because it includes the range of fundamental frequencies that are normally used in speaking and singing” (ibid.: p. 95). Laver also says (1980.: p. 110) that the modal voice should probably be differentiated into “chest voice” and “head voice”, but he says that he would use modal voice as a correspondent to “chest voice” register in his book. More on chest voice and head voice will be discussed in the paragraph 5. Laver (ibid.) also focuses on van den Berg’s (1968) work and how the modal voice is produced. Vocal cords are positioned in a manner which allows production of the modal voice. The main point is the lack of friction of the vocal cords.

The production of modal voice is thus carried out with only moderate adductive tension and moderate medial compression, with moderate longitudinal tension when the fundamental frequency is in the lower part of the range used in ordinary conversation. The vibration of the larynx in this condition is regularly periodic,

efficient in producing vibration, and without audible friction brought on by incomplete closure of the glottis. (Laver, 1980: p. 111)

3.1.3. Vocal fry

“Vocal fry, or creak phonation has the similarity with the neutral phonation in that it is also produced by vibrating vocal folds, however here a very low frequency is present being at around 25 to 50 Hz.” (Dogil, n.d.). In words of Laver (1980) vocal fry and harsh voice are quite similar. “The low fundamental frequency of this creak type of phonation is one factor that distinguishes it from harsh voice, which is otherwise somewhat similar.” (Laver, 1980: p. 122) Laver states that the mean fundamental frequency for vocal fry is 34.6 Hz and for harsh voice 122.1 Hz which is in the range of the modal voice. Laver (1980) also delved into Fónagy’s (1962) interpretation of the vocal fry and stated that the ventricular folds (placed right above the vocal cords) were pressed hard against each other, the ventricle of Morgagni (placed between the vocal cords and ventricular band) was wrinkled, the vocal folds were held tightly together and the air column went vertically through the larynx. The distinctive sound of the vocal fry is well described by Laver (1980: p. 124): “The effect of continual, separate taps in rapid sequence is an essential part of the characteristic auditory quality of creak.” These rapid taps are so fast that it almost sounds like a frog croaking. Vocal cry has been found in conversational speech of young adult, especially female, Standard American-English speakers (Wolk et al., 2012), and one of the most popular speakers with vocal fry is Kim Kardashian³. Vocal fry is used in many languages for different means, but in English, according to Laver (1980) it is used in Received Pronunciation in combination with falling intonation as means of showing the listener that you are ending your turn as a speaker as well as to show bored resignation during an utterance. Laver (ibid.) also states that it is sometimes believed that vocal fry overlaps with glottalization, but he says there is almost no important connection between the two so it should mostly be disregarded.

³<https://www.youtube.com/watch?v=R8mcBdBL-t0>

3.1.4. Breathy voice

“People with breathy voice have a low muscular tension, minimal adductive tension, weak medial compression and medium longitudinal tension of the vocal folds which results in the voice being silent and having the breathing a part of each word.” (Dogil, n.d.) Speakers with breathy voice exhale while talking and that breath can be heard simultaneously with words. Laver says about breathy voice and breathiness:

Breathiness is a quality which is quite often heard as a modification of modal voice (...) the mode of vibration of the vocal folds is inefficient, and is accompanied by slight audible friction. Muscular effort is low, with the result that the glottis is kept somewhat open along most of its length, and the folds never meet on the mid-line. Because each closing movement of the folds tends to be abortive, the lessened glottal resistance leads to a higher rate of air flow than in modal voice (Laver, 1980: p. 132)

What this means is that there is a lot of space for air to flow through the vocal folds, which leads to voice being produced while at the same time one is breathing out loudly. Breath and words are coming out at the same time which means it is difficult to continuously produce a breathy voice for a longer period of time. It is similar sounding to whispery voice “... there is a close auditory relationship between breathy voice and whispery voice (...) placing of the borderline between the two categories is merely an operational decision.” The main difference between the two is that whispery voice needs a higher medial compression than the breathy voice.” Laver (1980: p. 133). A good example for a speaker with a breathy voice is Marilyn Monroe.⁴

3.1.5. Harsh voice

On the opposite side of the spectrum are tense vocal folds being under constant pressure. This makes the voice sound somewhat angry and irritated, having therefore an appropriate name – harsh voice. The phonation is usually over 100 Hz (Dogil, n. d.). Laver (1980: p. 127) defines harsh voice with the following description: “The predominant characteristic

4

<https://www.youtube.com/watch?v=WkG40Yaz9fQ>

is the aperiodicity of the fundamental frequency, which is heard as a component of auditory quality rather than of auditory pitch.” Cooper, et. al. (1957: p. 183) called this aperiodicity of a pitch ‘jitter’, to which Laver (1980: p. 127) says that listeners are very sensitive about even small jitters which gives more insight to why harsh voice may sound rough and argumentative for the listeners. An example is Whoopi Goldberg.⁵

3.1.6. Falsetto

Finally, there is falsetto. Vocal folds become very thin after being stretched longitudinally. The vibration is not that considerable and the voice becomes high (Dogil, n. d.). Van Riper and Irwin (1958, as cited by Laver, 1980: p. 118) explain

The arytenoid cartilages adduct the vocal folds (...) The vocalis muscles along the glottal edge of each vocal fold remain relaxed, but the mass of each vocal fold is made stiff and immobile by contraction of the thyroarytenoid muscles, which make up the outer bulk of the folds. The vocal ligaments along the glottal edge of the vocal folds are put under strong tension (...) This results in the vertical cross-section of the edges of the vocal folds becoming thin. The glottis often remains slightly apart, and the characteristic sub-glottal air pressure is lower than for modal voice

As Laver (1980: p. 119) comments, there is a reduction of the expenditure of air in the vocal folds fry which gives it a church choir sound.

Some of these phonation types appear alone, while others can only appear in a combination with another. Laver (1980) states that modal voice and falsetto can appear by themselves but also in combination with other phonation types but not with each other. A good example would be Michael Jackson who used falsetto while singing.⁶

⁵<https://www.youtube.com/watch?v=GnFnFF7UI0g>

⁶https://www.youtube.com/watch?v=Q2JmWfXc1Ag&ab_channel=MichaelJacksonfan

3.2. Vocal quality combinations

According to Laver (1980: pp. 135-136) whisper and creak can appear alone or in combination with one another to form a whispery creak. They can also appear in a combination with phonation types from the first group to form whispery voice and whispery falsetto and creaky voice and creaky falsetto. Final combination is whispery creaky voice and whispery creaky falsetto.

Harshness and breathiness can only occur in combination with different phonation types, never alone. Harshness combines with modal voice and falsetto, creating harsh voice and harsh falsetto and breathiness combines with modal voice to give breathy voice. Multiple combinations of these phonation types are as follows: harsh whispery voice, harsh whispery falsetto, harsh creaky voice, harsh creaky falsetto, harsh whispery creaky voice, and harsh whispery creaky falsetto.

Understanding this helps us in decoding altered vocal qualities and personality stereotypes that we are going to discuss later on.

4. Vocal quality, paralanguage and cultural differences

Peng, et. al. (1993, as cited by Kreiman&Sidtis 2011: p. 353) conducted a study in which they discovered that people's perception of the speaker changes according to where the listeners come from. This occurs due to different cultures and learned beliefs, customs, and attitudes around the globe. People in Korea tend to have different expectations and affections towards a particular voice (ibid.). Peng, et. al. (1993, as cited by Kreiman&Sidtis, 2011: p. 353) found out:

...that Americans (who value youth above age) should hear faster or louder male voices as more dominant and competent, consistent with the impression of youth, while Koreans should hear louder or slower male talkers as more dominant and competent, consistent with the impression of age.

Furthermore, American listeners perceived the speakers who spoke more loudly and fast to be more powerful and competent as opposed to Koreans. They only had these conclusions about the loud talking speakers, meaning that Americans respect higher rate of speech as well as the tonality. It was also assumed (Kreiman&Sidtis, 2011) that the fundamental frequency (f0) differs across different nationalities. For instance, in Japan, female speakers tend to have higher pitched voices, which is considered to be more attractive in Japan than for instance, in the USA or Sweden. Kreiman&Sidtis (2011: p. 354) conducted a study in which they included female and male Dutch and Japanese listeners to listen to Japanese and Dutch female speakers and obtained similar results. The results showed that both of the groups found high-pitched female speakers to be less powerful, however, the results differed when the included factor for the listeners was the attractiveness of the voice.

Dutch listeners preferred medium- or low-pitched voices, while Japanese listeners preferred medium- or high-pitched voices, consistent with the hypothesis that Japanese women increase their speaking F0 to meet cultural standards for femininity. (Kreiman&Sidtis, 2011: p. 354)

This goes to show that people from different cultures perceive a certain voice and its characteristics differently. As per Kreiman&Sidtis (2011) the factors that influence this perception may be of sociocultural origin – female speakers in Japan tend to use higher voice to appear more polite and to fit the stereotypically female roles in Japanese culture. Graham et. al. (2001, as cited by Kreiman&Sidtis (2011: p. 304) proved that both Japanese and Spanish speakers had difficulty in

understanding emotions from English speakers, which was not the case with English listeners. However, when the listeners made errors in recognizing emotions, those errors that the Spanish speakers made were more similar to those of the English speakers than to those of the Japanese speakers, probably due to closer cultural correspondences (ibid.).

5. History of registers and vocal quality

5.1. History of registers

The analysis of the vocal apparatus and consequently vocal quality is very old. Hanson (1987, as cited by Finks, 1992: p. 3) said the following “The earliest identification of registers came from two medieval theorists named John of Garland (c. 1193-c. 1270) and Jerome of Moravia (c.1250). They believed the voice to have three registers.” According to Finks (1992) John of Garland wrote that the human voice exists in 3 forms, those being chest voice, throat voice and head voice. Garland meant for the chest voice to be a representation for low register, throat voice would be middle register and head voice was a symbol for high register, and singing registers would get a similar distribution. Finks (ibid.) also writes about Giulio Caccini (1546.-1618.) – a well-known singer, composer and a singing teacher. Caccini differentiated “voce piena” (real voice) from “voce finta” (distinguished voice). He taught his students to sing the low and the mid tones in the chest voice. The highest tones should be sung in falsetto. Caccini remains important to this day because of his teachings on blending the registers. There are more authors who analyzed vocalics and registers through music (Finks, 1992), but worthwhile mentioning is Pietro Francesco Tosi (1647.-1732.) who along with Giovanni Battista Mancini (1716.-1800.) identified two registers, naming them chest and head register. William Vennard (1967, as cited by Finks, 1992) discloses how there was some evidence from Tosi’s translations that Tosi also used a third register called falsetto but didn’t find it particularly useful. Miller (1987, as cited by Finks, 1992) deduces how it is very possible Tosi didn’t affirm the third register because Tosi was a *castrato*.

Castrato, also called Evirato, male soprano or contralto [is a] voice of great range, flexibility, and power, produced as a result of castration before puberty. The castrato voice was introduced in the 16th century, when women were banned from church choirs and the stage. (Britannica, n.d.)

A Spaniard by the name of Manuel Patricio Rodriguez Garcia, (1805.-1906.) a baritone and singing teacher, is credited for the invention of an early laryngoscope, according to Finks (1992). “...an instrument called the laryngoscope... This was nothing more than a small mirror fastened at the proper angle to a long handle. Until this invention there was no way to look down the throat.”

(ibid.: p. 8) It allowed people to get a new perspective on and more insight into an area previously unresearched. It allowed Rodriguez Garcia to understand how the larynx would move while singing, but it also showed the positioning of the tongue and the soft palate. This is the reason Garcia is marked as the first to emphasize the connection between laryngeal adjustment and vocal register (ibid.). Max Joseph Oertel (1835.-1897.) did a more thorough examination of the vocal cords with an instrument called laryngo-stroboscope (Finks, 1992) “The stroboscope is an instrument which enables us to see the intrinsic movements of any vibrating object.” (Finks, 1992: p. 11) However Oertel indicated there were only two registers, chest and upper register. In 1883 Emil Behnke and Lennox Browne differentiated between 5 registers. These were lower thick, upper thick, lower thin, upper thin and small and they built on Garcia’s 3 register theory. In 1905 Giovanni Battista Lamperti added another voice, namely medium voice, to Garcia’s chest and head voices. Lamperti was of the belief that too many voice doctors can ruin a singer’s prowess which is best showcased naturally. The example in which he described this was Garcia’s student Jenny Lind. At a young age, she lost her voice and tried regaining it while studying with Garcia. However, she failed but was later able to regain it herself and she became one of the best voices of her age according to Finks (1992). In 1950, a French scientist Husson (1901.-1967.) proposed that the neural impulses can do what air current usually does. It flows through the vocal cords and allows them to vibrate, creating voice. According to Finks (1992, p. 13) Husson claimed there were four registers. The “premier registre” or “register monophasé” or chest register, then the “second register” or “register biphasé” also called male falsetto or women’s head register, then there is the “troisième registre” or “register triphasé” which stands for the whistle register and finally the “quatrième registre” “register quadriphasé” is the one that correlates to the voice of an unusual singer (Finks, 1992, p. 13). In 1963, a Swedish voice teacher and scientist called Mörner (1895.-1977.) reviewed the terminology for different registers and composed a five register basis. Those five registers are: deepest range, deep level, mid level, high level, highest range (Finks, 1992, p. 14). In 1987 Johan Sundberg, (1936.-) the world’s foremost expert on the acoustics of singing put out a theory which differentiated male from female voices. In that regard, registers for the male voice were modal and falsetto and registers for the female voice were chest, medium and head according to Finks (1992: p. 16). To visualize the historic classification of registers and how it changed through the years, a table below will help.

Table 2 Terminology of registers (Finks, year, page)

Name	Preferred		Names of		Registers
John of Garland (1250)	Chest voice		Throat voice		Head voice
Jerome of Moravia (c. 1250)	Vox pectoris		Vox gutturis		Vox capitus
Caccini (1601)	Voice piena (naturale) natural voice		Voice finta feigned voice		
Cerone (1613)	Chest voice		Falsetto voice		
Tosi (1723)	Voice di petto		Voce di testa (falsetto)		
Mancini (1774)	Voice di petto		Voice di testa o falsetto		
Hiller (1774)	chest		Natural / head		falsetto
Martini (1792)	chest		throat		head
Manfredini (1775)	chest		Head / falsetto		
Müller (1837)	Chest		falsetto		
Garcia (1841)	Chest				Falsetto-head
(1856)	Chest		falsetto		Head
(1894)	Chest		Medium/falsetto (male)		Head
Oertel (1878)	Chest				Upper
Behnke and Bron (1883)	Thick lower thick/upper thick		Thin lower thin/upper thin		Small
MacKenzie (1885)	Long - reed		Short - reed		
G.B. Lamperti (1905)	chest		mixed		head
Nadoleczny (1923)	Not indicated				
Husson (1950)	Premier registre registremonophase		Second registre registrebiophase		Registretroisieme registretriphase
Oncley (1952)	Not indicated				
Van den Berg (1963)	Chest		mid		Falsetto/head
Mörner (1963)	Deepest range	Deep level	Mid level	High level	Highest range
Appelman (1967)	Chest		middle		Falsetto (male) head (female)
Large (1968)	chest		middle		Head/falsetto
Vennard (1967)	Heavy mechanism (chest)				Light mechanism (falsetto)
Fields (1970)	None				
Sundberg (1987)	Modal (male) chest (female)		Medium (female)		Falsetto (male) head (female)

The table above shows that across a 700 year span many authors used similar terms to describe different registers. Many of them used terms such as chest, head and middle voices, but they also

used different, more unique names. What is interesting is how all of them divided the voices in two or three fields. Even when more modern inventions helped them distinguish origins of the voices, they still used only two or three terms. Mörner was the only one who used five different terms for registers. However, it still fit the basic idea of the areas of vocal origin, only dividing what many others called chest area in two different areas - deepest range and deep level and also dividing what others called the middle area in two different areas - mid level and high level, with the third, which others called head register, being called highest range.

5.2. History of vocal quality

Vocal quality was taught all the way back in the antique times where it was important to be a good public speaker. In the ancient Greece there were different voice teachers with different roles according to Varošanec Škarić (2010: p. 15)

In the ancient Greece there were special voice masters (lat. edomatores vocis), which were divided into three groups: those who amplified the voice with exercise (lat. vociferarii), then those who tuned the voice quality, fullness, sonority, that is, they cultivated the pleasure of the voice (lat. phonascus) and final voice teachers who practiced the performance itself, developed expressiveness or voice expression and intonation (lat. vocales). Voice masters worked with vocal professionals, lawyers, actors, politicians. Phonetics experts and voice pedagogues later developed from the edomatores vocis⁷

According to Varošanec Škarić (ibid.) there were additional voice experts whose role was to take care of the health of the voice. Additionally, they assigned personal characteristics according to one's voice. Positive traits were associated with a deep and strong voice, which would imply that the speaker was brave. On the other side, high and weak voice carried negative characteristics, and had the traits of a scared person. Nasal voice was connected with morally weak speakers and hypocrites, bleating voice with dumb speakers, high, squeaky voice with envious speakers and cracked voices were reserved for bullies and dishonest people.

Along with the Greek scholars, Roman teachings were important, according to Varošanec Škarić (ibid.). Together with math, music, gymnastics, Roman children learned voice

⁷Originally in Coratian, translated by Jukić

exercises. Their politicians and rulers were also appreciated more if they had the ability to talk in a fine manner.

6. Personality stereotypes

Voice can reveal a lot about an individual. Many times, when we first hear someone speak, without seeing the speaker in person, we assume what they look like, what their personality is and most importantly – if we like the person or not – all based on the voice we hear.

Human brains make shortcuts. And with those shortcuts it can trick us into believing we know someone based on what others have shown us, or what we had learned prior to meeting them. A good example for this is a 2013 movie called “Iron Man 3”. In the movie, the main villain is called the Mandarin and he is presented as a dangerous terrorist threatening each day with a new attack. Here is the link⁸ to show his vocal characteristics. On top of his vocal characteristics, Mandarin uses words, which inflict fear on the listeners. These are words such as terrorist, attack, slaughter. In the video, it can also be seen how in the background of Mandarin’s speech, horrific scenes are played to instill even more panic. All of that makes for a panic inflicting scene which leaves common folk in fear of what is to come if it is not to be intervened. It is later revealed that the Mandarin was just a failed actor who got told what to say and how to say it by the actual movie villain. Every time people feared Mandarin, what they didn’t know was that he was a regular actor imitating a fearsome leader, and his performance was successful not only because of what he said, but of how he said it. The second link⁹, shows how the character (Trevor Slattery) actually speaks and should be a help to understand how the character changes his voice depending on the situation he is in. Opening the two links, two very different voices can be heard. The voice in the first link is deep, fearsome, and harsh. Voice is of very low frequency, the speaker is speaking slowly, and there is emphasis on many of his words, with additional pauses and a distinct accent, which brings chills to people listening, making him sound like a dangerous preacher. All of this contributes to make the message sound as dangerous as possible. In the second link we see and hear a complete opposite of the same person. Being ambushed and out of his role, Trevor Slattery shows his true nature. He is scared, his mind is influenced by drugs and he is speaking in a slightly high-pitched voice. He plays with words, using “he’s here but he’s not here” in a tensed voice and repeats the sentence a couple of times, showing that he is scared and that he is under stress. His voice becomes

⁸<https://www.youtube.com/watch?v=wDQquE4xxNc>

⁹<https://www.youtube.com/watch?v=IHPV7i-gfj4&t=149s>

falsetto and he is speaking at a faster rate showing signs of stress and anxiety. These clips show how a single man can change his vocal characteristics depending on the situation they find themselves in and the mood they are in.

7. Altered vocal quality

Vocal quality is specific for each individual. However, that doesn't necessarily mean that it may not be altered. This happens intentionally or unintentionally and may change the way the speaker is perceived. What this means is that people can have a different perspective of the speaker based on their voice. The listener may be attracted to a certain type of voice, or afraid of a different type of voice, resulting in the same feelings towards the speaker. This however, might not be the realistic representation of the speaker, as vocal quality does not always portray the personality of the speaker.

7.1. Medical conditions

Changed vocal quality can also happen due to different medical conditions or a psychological disorder. "... psychopaths described as "con artists," were found to speak more quietly and to use word emphasis differently than did non-psychopathic male offenders..." (Kreiman&Sidtis, 2011: p. 356) They do not differentiate emotional and non-emotional words which leads to them not understanding others as well as not understanding that what they are doing is wrong.

Another vocal altering factor is alcohol, which affects mood in low doses and as the concentration of alcohol in blood rises, so do the side effects. At higher rates, it affects motoric functions, walking capabilities and skilled operations. "With continued consumption, gait, speech, and other functions are further compromised, until coma develops at a blood alcohol concentration of about 0.3%." (Kreiman&Sidtis, 2011: p. 357) What interests us the most here is the compromised speech. It is not just the intonation of the sentences that get affected, but also the words and phrases. Sounds get substituted, e.g. /r/ becomes /l/, /s/ becomes /ʃ/ and /iz/ becomes /is/.

More subtle changes may also occur, including inappropriate lengthening of vowels and consonants, incomplete stop closure in affricates, deaffrication, and so on, leading to an overall impression of sloppy or imprecise pronunciation. The speaker may also have trouble maintaining the relative timing of different gestures needed to produce sounds, resulting in misordered speech sounds or production of the incorrect sound (Kreiman, Sidtis, 2011: p. 358).

This means that the speaker starts speaking more slowly and unintelligible. Words and sentences become harder to understand and even the meaning behind the sentence (or lack thereof) leads to even more pauses in the speech and eventually complete silence.

8. Vocal quality in modern times

Today, voice still has a strong role in the perception of the speaker. People use their voice not just for regular speaking but for effect. Stereotypes are what comes along with it. According to Hickson et. al. (2014) who analyzed Heinberg's (1964) theory which says that among 11 different voice types regarding pleasantness and unpleasantness, only one was perceived as pleasant. That one would be associated with proper breathing, articulation, tongue position, control of the pitch and resonance, etc. Hickson et. al. (ibid.) add that that voice would be the voice of a professional announcer. On the other side, Hickson et. al. (ibid.) inform on unpleasant voice types. These would be; breathy voice (females perceived as sexy and spacey; males as younger, artistic and effeminate), tense (older, anxious, nervous, uncooperative, less intelligent and high-strung), breathy-tense (four types, all of which are perceived as weak and nervous; husky, harsh, hard and strident), nasal (whiny, argumentative, lazy), denasal (stuffy, as if the speaker had a cold), orotund (energetic, pompous, authoritative, proud, humorless), flat (bored, sluggish, withdrawn), thin (immature, emotional, sensitive – mainly for females), throaty (sophisticated, less intelligent, careless, older), fronted (artificial, aloof).

It has also been found by Kramer (1978, as cited by Hickson et. al., 2014.) that male and female listeners agree in ideal speech types on 50 out of 51 characteristics identified. The listeners were high school and college students, among which males thought that the ideal voice would be a deep voice, whereas females disagreed. Much like being dressed poorly, or having a bad hairstyle, voice can push people away upon first meeting them. Listeners may not find the speaker attractive if the voice is not to their liking. "It appears that a "bad voice" can be a relationship disqualifier in a fashion similar to bad physical appearance." (Hickson et. al., 2014: p. 246). Hickson et. al. (ibid.) also write how according to a study made by Berry (1990, 1992) speakers with more attractive voices were considered more competent, warm and honest than people with unattractive voices, while people with babyish voices were found less competent and powerful, but more honest and warm, than those with mature voices.

Certain people have capabilities to recognize the speaker through their voice alone. This was exemplified in the movie Taken, where the main character recognizes a man by his voice, a clip of which is in the link here.¹⁰ However, in the recent past, voice has received a new purpose. Voiceprint is used in finding criminals through computer technology as a new wave of forensic

¹⁰<https://www.youtube.com/watch?v=xKQLVgrrvxl>

science. “The idea behind the voiceprint is that an individual’s voice is unique, much like a fingerprint, and that it is possible to identify people on the basis of certain vocal qualities.” (Hickson et. al., 2014: p. 248).

World is evolving rapidly over the years and with it, voices and perception of voices evolve as well. Much of it is influenced by technological advancements that have occurred in the recent years. Some 200, or even 50 years ago, technological breakthroughs that are today commonplace were unimaginable. One of those breakthroughs would be the synthetic voice. To have a better understanding of what synthetic voice actually is: “In many cases the synthesized voice is produced by concatenated segments of recorded human speech, a complex process that can result in an arguably lifeless voice, which lacks the ability for free-expression among other human qualities.” (Baird, et. al., 2018: p. 277). In other words, synthetic voice is voice which isn’t produced from a person’s throat, but from a machine. Synthetic voices still do have issues, they do sound fabricated, but they are getting better at it.

Voice-based interactions have previously been described as “frustrating and ineffective”, and in recent years large financial investments have been made by companies such as Amazon, Apple, Google, and IBM to improve both the design, and authenticity of the synthesized voice. (Baird, et. al.,2018: p. 277)

The reason it is difficult for a machine to fabricate a real voice is the lack of emotion and understanding of what is being said further than just plain words. For people to understand the message appropriately they need not only words, but paralinguistic cues as well. Without them, text messages can sometimes create confusion between the two speakers. According to Baird et al. (2018), it is also important what age and gender the speaker is. People use stereotypes to understand the speaker and his or her intentions. People may find a given compliment as either flirtatious or just friendly depending on the factors of age and gender. A machine can’t do that. It only receives information and fabricates an appropriate answer. The modern scientists are trying not only to replicate pitch and tone, but also non-linguistic features, such as face expressions, gestures and even body language as well.

The engineers behind the synthesized voice technology have been attempting to replicate the “muscular vibrations” produced by the body for many years, and such human-like features could also transfer to the machines they embody, making

understanding of such aspects an important factor for commercial success. (Baird, et.al., 2018: p. 278).

This means that in the future, machine fabricated speech could be almost identical to human speech and understanding it could be far easier, which would also make it more difficult to differentiate between the real human speech and the artificially fabricated speech.

Jekosch (2005) writes about the advancements in the developing systems which would be able to comprise a speaking pattern that would sound more authentic despite being machine made. To this day, it still sounds fabricated, it is easily recognizable as machine made and it has issues in understanding language produced by a real person. “However, it soon became apparent that natural speech sounds consist of far more than the concatenation of single speech sounds, syllables or words” (Jekosch, 2005: p. 28). Speech needs to have a better action and reaction activity which would include, as previously mentioned, paralinguistic features but also non-linguistic features. Making a machine laugh, or cry is still impossible or at least hard to accomplish for now. In order to understand what is being said, among other things, listener also needs to have some background knowledge of the speaker and should presume what the speaker might say. Human mind has that ability, while a computer still does not.

9. Conclusion

In conclusion, vocal quality is a trait which is present in every human being. By knowing more about it, we can understand each other better, but only if we learn more about its functions as well as human's ability to manipulate their voices. Majority of what and how we say something is happening without thinking, but there are instances where our speech can be manipulated and used to our advantage. If controlled, some parts of vocal quality can be altered and people's perception and reaction towards the speaker can be manipulated. By changing voice color, or depth, or talking speed, it is possible to create a different persona to the listeners.

Furthermore, this paper has also shown how stereotypes may distort the perception of the speaker and their discourse. Stereotypes are hard to dismiss once they are formed, but by educating the listener, one should get a less subjective perception of the speaker. What should not be discarded is how gender and age can affect both the speaker and the listener and these prejudices should also be worked on.

The history of vocal quality and works of different scholars have also been presented. Multiple authors throughout time have similarly named the registers, which shows that most of them had the same general idea on how a particular type of the voice is produced.

Finally, this paper has also shown some of the modern aspects on vocal quality. New advancements in technology are changing the perceptions about voice.

References

- 1 Minute of Kim Kardashian vocal fry - whang!. (2016). downloaded 10th of July 2022 from <https://www.youtube.com/watch?v=R8mcBdBL-t0>
- Baird A., Jørgensen H., Parada-Cabaleiro E., Cummins N., Hantke S., Schuller B., (2018). *The perception of Vocal Traits in Synthesized Voices: Age, Gender, and Human Likeness*. Augsburg, Copenhagen, München, London. ZD.B Chair of Embedded Intelligence for Health Care and Wellbeing, University of Augsburg, Germany 2 Department of Arts & Cultural Studies, University of Copenhagen, Denmark 3 MISP Group, MKK Technische Universität München, Germany 4 GLAM, Imperial College London, UK. p. 277-278
- Britannica thesaurus downloaded 2nd of July 2022 from <https://www.britannica.com/art/castrato>
- Cooper F.S., Peterson E., Fahringer G.S., (1957). *Some sources of characteristic vocoder quality*, *Journal of the Acoustical Society of America* 29. p. 183
- Esposito A., Faundez-Zanuy M., Keller E., Marinaro M. (2007). *Verbal and nonverbal communication behaviours*. Berlin. Springer-Verlag
- Finks P. L., (1992). *Vocal Registration: History, Analysis, and Modern Pedagogical Applications*. Illinois. Eastern Illinois University
- Hickson M. III, Moore N-J., Stacks D. W., (2014). *Nonverbal communication*. Oxford. Oxford University Press
- Hollien H., (1971). *Three major vocal registers: a proposal*, *Proceedings of the 7th International Congress of Phonetic Sciences*. Montreal.
- Jekosch U., (2005). *Voice and Speech Quality Perception*. Berlin. Springer Berlin Heidelberg
- Kreiman J., Van Lancker Sidtis D., (2011). *Perception of Emotion and Personality from Voice*. Wiley Blackwell
- Kreiman J., Van Lancker Sidtis D., Gerratt B. (2011). *Defining and measuring voice quality*. downloaded 5th of July 2022 from

https://www.researchgate.net/publication/237236417_Defining_and_measuring_voice_quality

Ladefoged, P. (1971) *Preliminaries to linguistic phonetics*. Chicago: University of Chicago

Laver J (1980). *The phonetic description of voice quality*. Cambridge. Cambridge studies in linguistics

Legati I (2008). *Zloporaba govorničke tehnike u javnom govorenju [Abuse of speaking technique in public speaking]*. Zagreb. Akademija dramske umjetnosti

Marilyn Monroe talking. (2021). Downloaded 30th of August 2022 from <https://www.youtube.com/watch?v=WkG40Yaz9fQ>

Michael Jackson's amazing falsetto (2015) Downloaded 18th of August 2022 from <https://www.youtube.com/watch?v=Q2JmWfXc1Ag>

Poyatos F. (1993). *Paralanguage: A linguistic and interdisciplinary approach to interactive speech and sounds*. Amsterdam/Philadelphia. John Benjamin's publishing company

Prejudice is a basic human need. (2011). downloaded 6th of July 2022 from <https://www.psychologicalscience.org/news/prejudice-is-a-basic-human-need.html>

Taken (2008) – “I Told You I Would Find You” (2012) Downloaded 22nd of August 2022 from <https://www.youtube.com/watch?v=xKQLVgrrvXI>

The Mandarin's History Lesson - Iron Man 3. (2014). downloaded 3rd of August 2021 from <https://www.youtube.com/watch?v=wDQquE4xxNc>

Tony Stark Meets Fake Mandarin Trevor Slattery Iron Man 3 2013. (2020). downloaded 3rd of August 2021 from <https://www.youtube.com/watch?v=IHPV7i-gfj4&t=149s>

Van Riper C., Irwin J. V., (1958) *Voice and articulation*, Prentice-Hall, Englewood Cliffs

Varošaneć Škarić G. (2010). *Fonetska njega glasa i izgovora [Fonetical care of voice and pronunciation]*. Zagreb. Filozofski fakultet

Whoopi Goldberg- Interview (Ghost) 7-1-90 [Reelin' In The Years Archive]. (2021).
Downloaded 30th of August 2022 from
<https://www.youtube.com/watch?v=GnFnFF7U10g>

Wolk, L., Abdell-Beruh, N. B., Slavin, D. (2012). *Habitual Use of Vocal Fry in Young Adult Female Speakers*. *Journal of Voice*, 26 (3), pp. 111-116.
doi:10.1016/j.jvoice.2011.04.007